Visually Lossless Coding for the HEVC Standard: Efficient Perceptual Quantisation Contributions for HEVC

by

Lee James Prangnell
MSc, University of Liverpool, 2012

A THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF:

Doctor of Philosophy

in

Computer Science

FACULTY OF SCIENCE
THE UNIVERSITY OF WARWICK

September 2017

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed solely by myself and has not been submitted in any previous application for any degree. I carried out all of the work including the scientific research, the novel technique development, the design of the experimental evaluations, the data analysis, the written work and the proofreading.

All of the contributions proposed in the thesis are based on several publications; the publications are listed below. Note that the co-authors' contributions in each paper consisted of providing useful feedback in order to strengthen the publication. I carried out all of the work including the scientific research, the novel technique development, the design of the experimental evaluations, the data analysis, the written work and the proofreading.

L. Prangnell, "Visually Lossless Coding in HEVC: A High Bit Depth and 4:4:4 Capable JND-Based Perceptual Quantisation Technique for HEVC," *Elsevier Signal Processing: Image Communication Journal*, vol. 63, pp. 125-140, 2018.

L. Prangnell and V. Sanchez, "JND-Based Perceptual Video Coding for 4:4:4 Screen Content Data in HEVC," in *Proceedings of the 43rd IEEE International Conference on Acoustics, Speech and Signal Processing*, Calgary, Alberta, Canada, 2018. In Press.

L. Prangnell, M. Hernández-Cabronero and V. Sanchez, "Coding Block Level Perceptual Video Coding for 4:4:4 Data in HEVC," *in Proceedings of the 24th IEEE International Conference Image Processing,* Beijing, China, 2017, pp. 2488-2492.

L. Prangnell, M. Hernández-Cabronero and V. Sanchez, "Cross-Color Channel Perceptually Adaptive Quantization for HEVC," in *Proceedings of the 27th IEEE Data Compression Conference*, Snowbird, Utah, USA, 2017, pp. 456.

L. Prangnell, V. Sanchez and R. Vanam, "Adaptive Quantization by Soft Thresholding in HEVC," in *Proceedings of the 31st IEEE Picture Coding Symposium*, Cairns, Queensland, Australia, 2015, pp. 35-39.

Acknowledgements

I dedicate this PhD thesis to my beloved mother and father. The unconditional love, the outstanding emotional support and also the wonderful encouragement they have consistently provided during my candidacy have undoubtedly facilitated my academic progression, for which I am — and always will be — eternally grateful.

The research in this thesis was conducted in the SIP Lab within the Multimedia Processing and Computer Vision Laboratory. I received a full PhD scholarship from The Engineering and Physical Sciences Research Council (EPSRC) to fund the PhD research.

I would like to thank all staff in the Department of Computer Science. In particular, I wish to thank my supervisor, Dr. Victor Sanchez, for the excellent supervision, feedback and advice that he has consistently provided during the PhD course. With respect to the annual review procedure and related duties, I am grateful to Prof. Chang-Tsun Li for acting as my PhD advisor. In addition, I wish to show gratitude to Dr. Alexandra Cristea and also Sharon Howard for the all of the help and support that they provided during my PhD registration.

In terms of collaborations, I wish to thank to Dr. Victor Sanchez, Dr. Miguel Hernández-Cabronero and Dr. Rahul Vanam for their valuable feedback as co-authors of my conference publications. In addition, I wish to thank Shevach Riabtsev for the useful feedback that he provided for my journal publication.

Last but not least, I am extremely grateful to Rebecca Fairchild — formerly of Warwick Mental Health Team — for the exceptional advice and support that she consistently provided in relation to the psychiatric illnesses with which I am afflicted.

Abbreviations and Acronyms

1080p	1920×1080 Pixel Resolution
4K	3840×2160 Pixel Resolution (2160p)
8K	7680×4320 Pixel Resolution (4320p)
AC	Alternating Current
AI	All Intra
AIP	Angular Intra Prediction
AVC	Advanced Video Coding Standard (H.264/MPEG-4 Part 10)
AVMP	Advanced Motion Vector Prediction
CABAC	Context Adaptive Binary Arithmetic Coding
Cb	Chroma (Blue Difference)
CBPQ	Coding Block Perceptual Quantisation
CCCPQ	Cross-Colour Channel Perceptual Quantisation
Cr	Chroma (Red Difference)
CRT	Cathode Ray Tube
CSF	Contrast Sensitivity Function
CB	Coding Block
CPU	Central Processing Unit
CT	Computerised Tomography
CTB	Coding Tree Block
CTU	Coding Tree Unit
CU	Coding Unit
DC	Direct Current
DCT	Discrete Cosine Transform
DDR3	Double Data Rate (Type Three)

DDR5	Double Data Rate (Type Five)
DF	Deblocking Filter
DST	Discrete Sine Transform
DRAM	Dynamic Random Access Memory
eV	Electronvolts
FWM	Frequency Weighting Matrix
GB	Gigabytes
GHz	Gigahertz
GOP	Group of Pictures
GPU	Graphics Processing Unit
HD	High Definition
HDR	High Dynamic Range
HM	JCT-VC HEVC Reference Software
HEVC	High Efficiency Video Coding Standard
HVS	Human Visual System
IDSQ	Intensity Dependent Spatial Quantisation
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardisation
ITU	International Telecommunication Union (United Nations)
J	Joules
JCT-VC	Joint Collaborative Team on Video Coding
JND	Just Noticeable Distortion
JPEG	Joint Photographic Experts Group
Kbps	Kilobits Per Second

L, M, S	Long, Medium, Short
LCD	Liquid Crystal Display
LCU	Largest Coding Unit
LED	Light Emitting Diode
m/s	Metres Per Second
MB	Megabytes
MC	Motion Compensation
MDIS	Mode Dependent Intra Smoothing
ME	Motion Estimation
MF	Multiplication Factor
MHz	Megahertz
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MSE	Mean Squared Error
nm	Nanometres
PB	Prediction Block
PET	Positron Emission Tomography
PhD	Doctor of Philosophy Degree
Pixel-PAQ	Pixel Perceptually Adaptive Quantisation
PSNR	Peak Signal-to-Noise Ratio
PU	Prediction Unit
QM	Quantisation Matrix
QMs	Quantisation Matrices
QP	Quantisation Parameter
QStep	Quantisation Step Size

QT	QuadTree
RA	Random Access
RDO	Rate Distortion Optimisation
RDOQ	Rate Distortion Optimised Quantisation
RDT	Rate Distortion Theory
RGB	Red, Green, Blue (Colour Space)
RSP	Reference Sample Padding
SAO	Sample Adaptive Offset
SCU	Smallest Coding Unit
SDRAM	Synchronous DRAM
SF	Scaling Factor
SSIM	Structural Similarity Index
TB	Transform Block
TCPQ	Transform Coefficient Perceptual Quantisation
THz	Terahertz
TU	Transform Unit
TV	Television
URQ	Uniform Reconstruction Quantisation
VCEG	Video Coding Experts Group
VDU	Visual Display Unit
WCG	Wide Colour Gamut
Y	Luma
YCbCr	Luma and Chroma Colour Space (Digital)
YUV	Luma and Chroma Colour Space (Analogue)

Abstract

In the context of video compression, visually lossless coding refers to a form of perceptual compression. The objectives are as follows: i) to lossy code a raw video sequence to the lowest possible bitrate; ii) to ensure that the compressed sequence is perceptually identical to the raw video data. Because of the vast bitrate reductions which cannot otherwise be achieved, the research and development of visually lossless coding techniques (e.g., perceptual quantisation methods) is considered to be important in contemporary video compression research, particularly for the High Efficiency Video Coding (HEVC) standard.

The default quantisation techniques in HEVC — namely, Uniform Reconstruction Quantisation (URQ) and Rate Distortion Optimised Quantisation (RDOQ) — are not perceptually optimised. Neither URQ nor RDOQ take into account the Modulation Transfer Function (MTF)-based visual masking properties of the Human Visual System (HVS); e.g., luma and chroma spatial masking. Moreover, URQ and RDOQ do not intrinsically possess the capacity to distinguish luma data from chroma data. Both of these shortcomings can lead to coding inefficiency (i.e., wasting bits by not removing perceptually irrelevant data). Therefore, it is desirable to develop visually lossless coding (perceptual quantisation) techniques for HEVC. For example, by taking chrominance masking into account, perceptual quantisation techniques can be designed to discard — to a very high degree — chroma-based psychovisual redundancies from the chroma channels in raw YCbCr video data. To this end, four novel perceptual quantisation contributions are proposed in this thesis.

In Chapter 3, a novel transform coefficient-level perceptual quantisation method is proposed. In HEVC, each frequency sub-band in the Discrete Cosine Transform (DCT) frequency domain constitutes a different level of perceptual importance to the HVS. In terms of perceptual importance, the DC coefficient (very low frequency) is the most important transform coefficient, whereas the AC coefficients farthest away from the DC coefficient (very high frequency AC coefficients) are the least perceptually relevant. Therefore, the proposed technique is designed to quantise AC coefficients based on their Euclidean distance from the DC coefficient.

In Chapter 4, two novel perceptual quantisation methods are proposed, which are based on HVS visual masking in the spatial domain. The first technique operates at the Coding Unit (CU) level and the second operates at the Coding Block (CB) level. Both techniques exploit the fact that the HVS can tolerate high levels of distortion in high variance (busy) regions of compressed luma and chroma data. The CU-level method adjusts the Quantisation Parameter (QP) of a $2N \times 2N$ CU based on cross colour channel variance computations. The CB-level technique separately adjusts the QP of the Y, Cb and Cr CBs in a CU based on separate variance computations in each colour channel.

In Chapter 5, a novel CB-level luma and chroma perceptual quantisation technique — based on a Just Noticeable Distortion (JND) model — is proposed for HEVC. The objective of this technique is to attain visually lossless coding at extremely low bitrates by exploiting HVS-related luminance adaptation and chrominance adaptation. Consequently, this facilitates JND perceptual quantisation based on luminance spatial masking and chrominance spatial masking. The proposed technique applies high levels of perceptual quantisation to luma and chroma data, which is achieved by separately adjusting the Quantisation Step Sizes (QSteps) at the level of the Y CB, the Cb CB and the Cr CB in a CU. To the best of the author's knowledge, this is the first JND-based perceptual quantisation technique that is compatible with high bit depth YCbCr data irrespective of its chroma sampling ratio.

The novel techniques proposed in this thesis are evaluated thoroughly. The methodology utilised in the experiments consists of an exhaustive subjective visual quality assessment in addition to an extensive objective visual quality evaluation. The subjective evaluation is based on the International Telecommunications Union (ITU) standardised assessments known as ITU-R: Rec. P.910. In these tests, several participants undertake a considerable number of subjective visual inspections (e.g., spatiotemporal analyses of the compressed sequences versus the raw video data) to ascertain the efficacy of the proposed contributions. The objective visual quality evaluation includes quantifying the mathematical reconstruction quality of the video data compressed by the proposed techniques. This is carried out by employing the Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) visual quality metrics.

Chapter 1. Introduction

Raw video data in the YCbCr colour space contains various forms of redundancy in each colour channel including spatial, temporal and psychovisual redundancies [1, 2]. A raw YCbCr sequence typically requires a vast bandwidth capacity. The bitrate of a sequence is dictated by its pixel resolution, its bit depth, its chroma sampling ratio and the number of frames in the sequence. Using the example of a raw 1080p 10-bit YCbCr (4:4:4 full chroma) sequence comprising 150,000 frames (e.g., a two hour sequence), this type of sequence could easily consume many megabits per second, which is impractical. This is the reason why video compression is necessary.

In terms of compressing raw video data while maximally preserving its fidelity in HEVC, this can be achieved by employing mathematically lossless compression (i.e., no compression distortion: a Mean Squared Error (MSE) = 0 and a PSNR = ∞ dB). However, the bitrate reduction attained by mathematically lossless compression is usually very modest. Therefore, lossy coding dominates contemporary video compression research because of the comparatively high bitrate reductions it attains.

Lossy coding algorithms in video compression standards employ Rate Distortion Optimisation (RDO), which is based on Rate Distortion Theory (RDT [3]). In HEVC [4, 5], spatial and temporal redundancies are discarded from raw video data using RDO-based lossy coding algorithms [6]. RDO facilitates an acceptable balance between bitrate reduction and quantisation-induced distortion. However, RDO-based lossy coding techniques do not maximally discard psychovisual redundancies from raw video data, which has given rise to visually lossless coding research.

Visually lossless coding methods, such as perceptual quantisation, are designed to facilitate a more comprehensive removal of perceptually irrelevant psychovisual redundancies from raw video data (e.g., imperceptible colour differences by virtue of chrominance masking). To this end, perceptual quantisation techniques in HEVC can augment RDO-based lossy coding techniques, thereby further reducing bitrates to a considerable degree. For this reason, visually lossless coding is presently of great interest to the video compression research community [2].

1.1 Overview of the HEVC Standard and Quantisation in HEVC

The Joint Collaborative Team on Video Coding (JCT-VC) developed and standardised HEVC [4, 5] to supersede the ubiquitous Advanced Video Coding (AVC) standard [7, 8]. The standardisation of HEVC version 1 took place in January 2013; note that HEVC version 5 is the latest version of the standard [4, 5]. In comparison with AVC, the key improvement that HEVC attains is the outstanding coding efficiency improvement that it yields. With its enhanced video coding algorithms, HEVC improves coding efficiency by up to 50% compared with AVC. These vast improvements in coding efficiency facilitate the coding and decoding of high quality bitstreams for utilisation on appreciably high display resolution environments (including Ultra HD 4K and 8K).

In basic terms, the video encoder in HEVC involves the following processes. Firstly, the raw video data is partitioned into blocks. After this, HEVC processes the partitioned data using algorithms known as intra prediction (spatial prediction) and inter prediction (spatiotemporal prediction). Intra prediction errors [9] and inter prediction errors [10, 11] produce luma and chroma (Cb and Cr) residual values. Subsequently, integer approximations of the DCT and the Discrete Sine Transform (DST) convert the prediction residual values into the frequency domain [12, 13], which produces transform coefficients. Quantisation of the transform coefficients then takes place using a combination of URQ and RDOQ. Finally, Context Adaptive Binary Arithmetic Coding (CABAC) — which is a lossless compression algorithm — compresses the quantised transform coefficients.

The block partitioning structure in HEVC includes Coding Tree Units (CTUs), Coding Units (CUs), Prediction Units (PUs) and Transform Units (TUs) [14]. Initially, the raw video data is partitioned into CTUs (see Figure 2.2); the CTU is the fundamental logical unit in HEVC. The CTU can be considered as a set of four constituent components (CTU: the Y Coding Tree Block (CTB), Cb CTB and Cr CTB constituents and the associated syntax elements); the same is true for CUs, PUs and TUs. Note that each CTB is 16×16 to 64×64 samples in size.

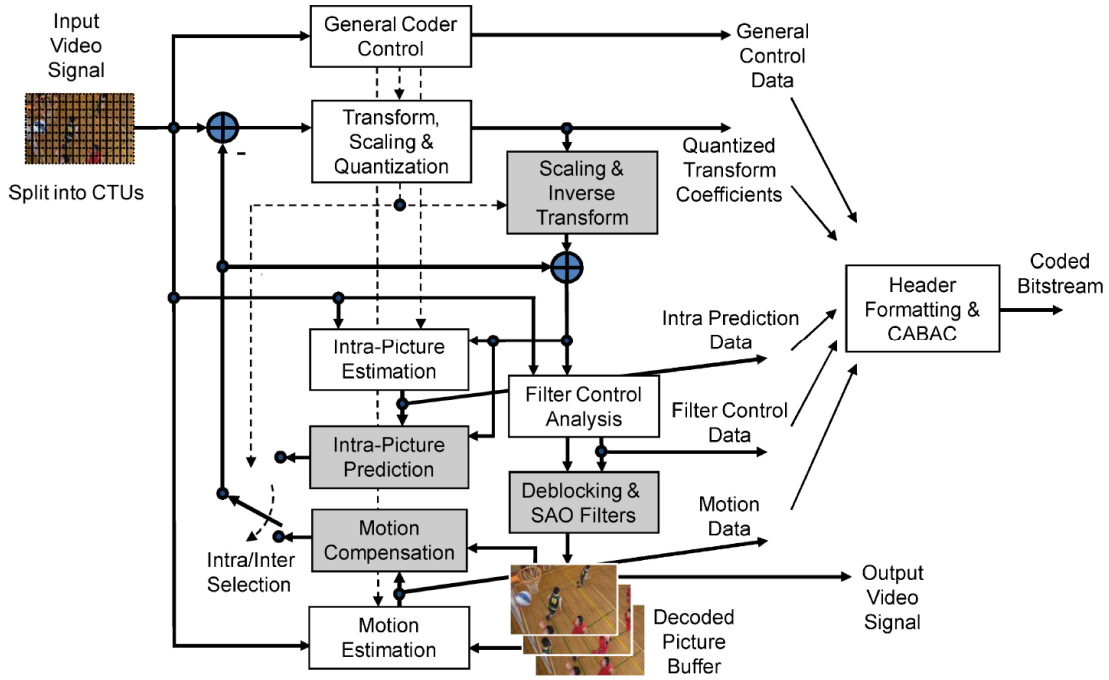


Figure 2.2: A block diagram, developed by Sullivan et al. [5], of a typical video encoder in the HEVC standard.

The partitioning mechanism in HEVC splits a CTU into four constituent CUs for the prediction decision process. The CU assumes the role of the decision making point in terms of choosing whether to perform either intra prediction only or combined intra prediction and inter prediction. The Largest Coding Unit (LCU) is 64×64 samples in size and the Smallest Coding Unit (SCU) is 8×8 samples in size [14]. The CU, in which the luma and chroma CBs and syntax elements reside, contains all of the prediction information required for the coding process. It is important to note that the size of the chroma Cb and Cr CBs in each CU depends on the chroma sampling ratio in the raw YCbCr video data [15, 14].

The decision to code an area of a frame — using intra prediction and/or inter prediction — is made at the CU level; however, a CU can be split into a PU depending on the prediction type decision. For example, an 8×8 luma CB within a CU may be too large to store important luma inter prediction information, such as a motion vector with a very small magnitude. As implied previously, a PU contains a Y Prediction Block (PB), a Cb PB and a Cr PB in addition to the associated syntax elements; each PU is 4×4 to 64×64 samples in size.

Like the PU partitioning structure, the TU block transform structure has its root at the CU level. As stated in [4, 5], a CB-level luma prediction residual block can be identical to a luma TB. Furthermore, the residual block can be further partitioned into smaller TBs depending on the type of residual block, the energy of the residual block and the type of linear transformation that is applied to the residual values (e.g., the integer approximation of DCT). Note that the sizes of the luma and chroma TBs range from 4×4 to 32×32 samples; a larger transform typically equates to improved coding efficiency (at the expense of increased computational complexity) [16].

In HEVC, there is a sophisticated form of intra prediction, which includes angular intra prediction (33 spatial prediction modes for luma and chroma data), mode dependent intra smoothing and reference sample padding [9]. All Intra (AI) coding (i.e., coding using intra prediction only) is seldom utilised in contemporary lossy video compression applications. This is because of the high bitrate costs incurred due to the lack of Group Of Pictures (GOP)-based inter-frame considerations. Therefore, if All Intra coding mode is employed, quantisation-induced compression artifacts are typically vastly more conspicuous in the compressed video data. In lossy video coding, the useful combination of intra prediction and inter prediction is ubiquitous. In this mode, intra coded frames are utilised as key frames, after which bidirectional inter prediction — e.g., using the Random Access (RA) configuration — is utilised for subsequent frames. In HEVC, inter prediction includes motion estimation, motion compensation and advanced motion vector prediction [10, 11]. Motion data coded with GOP-based inter prediction can be signalled using merge mode, motion vector differences, picture reference indices and the direction of the inter prediction [10].

As previously mentioned, HEVC includes finite precision integer approximations of the DCT and the DST [12, 13]. These techniques transform the intra prediction and inter prediction residual data from the spatial domain into the frequency domain. Recall that the DC transform coefficient and the low frequency AC transform coefficients contain the most important energy in terms of how the HVS perceives the compressed data. To reiterate, quantisation is primarily designed to discard (zero out) high frequency AC coefficients. In addition, coefficients in the medium frequency sub-band can also be zeroed out; this depends on the level of quantisation specified, which is determined by the QP and Quantisation Step Size (QStep) [17].

URQ is the default uniform quantisation technique in HEVC [5, 13]; it quantises transformed luma and chroma residual values (i.e., transform coefficients). URQ includes a QStep value — controlled by the QP — that applies equal levels of quantisation to all transform coefficients in luma and chroma TBs irrespective of the frequency sub-band of a coefficient. Therefore, this means that URQ is not perceptually optimised, thus leaving room for improvement. The QStep in URQ increases uniformly by approximately 12% for each value increment of the QP [17]. Note that there is a binary logarithmic relationship between the QP and the QStep. The luma and chroma QSteps, denoted as $QStep_Y$, $QStep_{Cb}$ and $QStep_{Cr}$, are defined in (1.1) to (1.3), respectively:

$$QStep_Y\left(QP_Y\right) = 2^{\frac{QP_Y-4}{6}} \tag{1.1}$$

$$QStep_{Cb}\left(QP_{Cb}\right) = 2^{\frac{QP_{Cb}-4}{6}} \tag{1.2}$$

$$QStep_{Cr}\left(QP_{Cr}\right) = 2^{\frac{QP_{Cr}-4}{6}} \tag{1.3}$$

where QP_Y, QP_{Cb} and QP_{Cr} correspond to the integer luma and chroma QP values, respectively; they are defined in (1.4) to (1.6), respectively.

$$QP_Y\left(QStep_Y\right) = \left\lfloor 6 \times \log_2\left(QStep_Y\right)\right\rfloor + 4 \tag{1.4}$$

$$QP_{Cb}\left(QStep_{Cb}\right) = \left\lfloor 6 \times \log_2\left(QStep_{Cb}\right)\right\rfloor + 4 \tag{1.5}$$

$$QP_{Cr}\left(QStep_{Cr}\right) = \left\lfloor 6 \times \log_2\left(QStep_{Cr}\right)\right\rfloor + 4 \tag{1.6}$$

Note that URQ cannot distinguish luma coefficients from chroma coefficients, which constitutes another shortcoming. However, JCT-VC has provided the flexibility of signalling chroma QP offsets, with respect to the luma QP, at both the frame level and the CB level [18, 19], which facilitates potential modifications to URQ in this regard.

To reiterate, after quantisation, CABAC losslessly compresses the quantised transform coefficients; this is the point at which the actual data compression takes place in HEVC. Assuming that high levels of quantisation have been applied to coefficients during the coding process (i.e., a high QP value has been applied), this gives rise to a decrease in non-zero quantised coefficients. CABAC can then efficiently quantise the zeroed out coefficients, thus resulting in a compressed bitstream with fewer bits.

1.1.1 Visually Lossless Coding (Perceptual Quantisation) for HEVC

In HEVC, the DCT and DST basis functions correspond to the Modulation Transfer Function (MTF) characteristics of the HVS [12, 13]; the MTF is derived from the Contrast Sensitivity Function (CSF). Focusing on the DCT, it prioritises the preservation of low frequency data. This is because the HVS is vastly more sensitive to gradations in the low frequencies of compressed video and image data (e.g., quantisation-induced compressed artifacts). The DCT linear transformation compacts the prediction residual energy into the DC coefficient and the very low frequency AC coefficients. In terms of scalar quantisation in HEVC, it is desirable to quantise perceptually insignificant AC coefficients more aggressively to maximise bitrate reductions. URQ does not take into account the perceptual relevance of individual transform coefficients at different frequency sub-bands. Therefore, URQ does not take advantage of the MTF characteristics of the HVS, which leaves significant room for improvement. In this context, the visually lossless coding (perceptual quantisation) technique proposed in Chapter 3 [21] significantly improves upon URQ.

Recall that URQ applies equal levels of quantisation to coefficients in both luma and chroma TBs. This means that URQ cannot distinguish the difference between luma coefficients and chroma coefficients, which constitutes another shortcoming. The HVS is considerably more sensitive to compression artifacts in achromatic data. Conversely, the HVS can tolerate much higher degrees of compression-induced noise in chrominance data. Therefore, applying equal levels of quantisation to both luma and chroma coefficients, as is the case with URQ, is not perceptually optimal; chroma data can be quantised to a much higher degree than luma data. The techniques proposed in Chapters 4 and 5 ([22, 23] and [24, 25], respectively) exploit this fact.

1.2 Experimental Evaluation Procedure

Before providing an overview of the proposed four perceptual quantisation contributions, we firstly provide detailed expositions of the experimental setup. An exhaustive evaluation procedure is utilised on all four techniques proposed in this thesis (i.e., the novel techniques proposed in Chapter 3, Chapter 4 and Chapter 5). The objective visual quality evaluations correspond, as closely as possible, to the Common Test Conditions and Software Reference Configurations recommended by JCT-VC [26, 27]. The experimental setup utilised in this thesis includes testing the proposed techniques with five initial QP data points (i.e., QPs 17, 22, 27, 32 and 37) and the All Intra (AI) and Random Access (RA) encoding configurations [26, 27].

In the objective visual quality evaluations, we employ the Structural Similarity Index (SSIM) [28] and the PNSR visual quality metrics to assess the reconstruction quality of the compressed video frames; i.e., in the sequences coded by the proposed techniques and the reference techniques (anchors). The SSIM and PSNR values presented in each contribution chapter correspond to objective spatial reconstruction analyses of the intra-frames (in the AI tests) and also the inter-frames — in addition to the intra key frames — (in the RA tests), per sequence. In other words, the mean SSIM and PSNR values are computed by comparing the coded frames, using initial QPs 17, 22, 27, 32 and 37, with the frames in the raw data; this is carried out in bitmap image form for each sequence and each QP.

Each novel technique proposed in this thesis is implemented into the JCT-VC HEVC HM 16.7 reference software [29, 30]; they are evaluated on 18 official test sequences provided by JCT-VC. The test sequences are as follows: the YCbCr 4:2:0, 4:2:2 and 4:4:4 versions of the BirdsInCage, DuckAndLegs, Kimono, OldTownCross, ParkScene and Traffic sequences (note that the CBPQ technique in Chapter 4 is evaluated on the 4:4:4 version only). Figures 1.1 - 1.6 show a frame from each 4:4:4 raw sequence in bitmap image form. All of the aforementioned sequences comprise a spatial resolution of full High Definition (HD), 1920×1080 pixels (1080p). The 4:4:4 and 4:2:2 versions of these sequences contain a higher dynamic range (i.e., 10-bits per sample per channel, which equates to 30-bits per sample), whereas the 4:2:0 versions comprise 8-bits per sample per channel (24-bits per sample).

Figure 1.1: JCT-VC BirdsInCage 4:4:4 Raw Data (HD 1080p) for HEVC Evaluations

Figure 1.2: JCT-VC DuckAndLegs 4:4:4 Raw Data (HD 1080p) for HEVC Evaluations

Figure 1.3: JCT-VC Kimono 4:4:4 Raw Data (HD 1080p) for HEVC Evaluations

Figure 1.4: JCT-VC OldTownCross 4:4:4 Raw Data (HD 1080p) for HEVC Evaluations

Figure 1.5: JCT-VC ParkScene 4:4:4 Raw Data (HD 1080p) for HEVC Evaluations

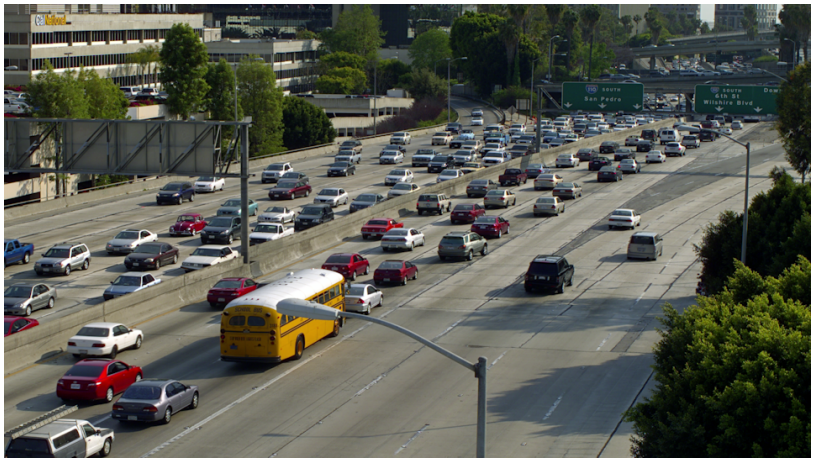


Figure 1.6: JCT-VC Traffic 4:4:4 Raw Data (HD 1080p) for HEVC Evaluations

Table 1.1. The criteria for quantifying the MOS with respect to the visual reconstruction quality of a compressed video sequence (compared with the raw video data).

MOS	Visual Quality Difference
5	Imperceptible (Visually Lossless)
4	Very Slightly Perceptible
3	Moderately Perceptible
2	Significantly Perceptible
1	Extremely Obvious

Because all of the proposed techniques are optimised for perceptual compression, it is of significant importance to undertake exhaustive subjective visual quality evaluations (in addition to the aforementioned objective visual quality evaluations). Therefore, we employ a United Nations' ITU-T standardised subjective evaluation procedure [31]. The subjective visual quality evaluations are the most important set of experiments in terms of measuring the visually quality of perceptually compressed video sequences (especially so for extremely low bitrate visually lossless coding techniques).

The computer hardware utilised in the experimental setup consists of a desktop PC that contains an Intel Core i7-4770 CPU — 4 cores and 8 threads — running at 3.4 GHz per core. The volatile memory in the PC is as follows: 24 GB of Double Data Rate type 3 (DDR3) Synchronous Dynamic Random Access Memory (SDRAM) running at a frequency of 680 MHz. The GPU installed in the PC is an NVIDIA GeForce 750 Ti with a core clock speed of 1020 MHz, 2 GB of DDR5 SDRAM and a memory bandwidth of 86.4 GHz/s. It is an Ultra HD 4K and High Dynamic Range (HDR)-capable GPU that can support YCbCr 4:4:4 and RGB data of bit depths up to 12-bits per sample per channel (i.e., 36-bits per sample). The subjective evaluations are conducted on the following TV/Visual Display Unit (VDU): HD 1080p 32 inch Samsung F5500 LED Smart TV. On the aforementioned hardware used in the experimental setup, note that, due to the higher dynamic ranges and the increased levels of chroma saturation in the raw YCbCr 4:4:4 and 4:2:2 10-bit sequences, the superior visual quality of this data is perceptually discernible in comparison with the chroma subsampled raw 4:2:0 8-bit sequences.

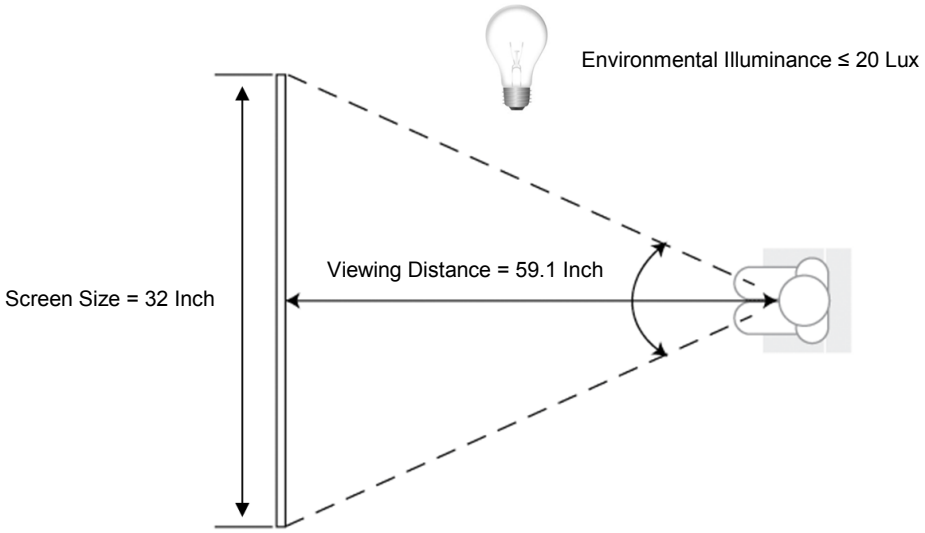


Figure 1.7: A diagram that illustrates the viewing conditions of the subjective evaluations. This includes the screen size of the TV/VDU, the viewing distance of the participant from the TV/VDU and the illuminance.

The main objectives of the subjective evaluations are as follows: i) to establish if the proposed techniques achieve visually lossless coding — compared with the raw video data — at low bitrates; ii) to ascertain if the proposed techniques outperform or match the performance of the references techniques (anchors) in terms of visual quality. To reiterate, we follow, as closely as possible, the directions of the internationally standardised United Nations' ITU subjective evaluations entitled: *Subjective Video Quality Assessment Methods* (ITU-R Rec. P.910 [31]). In ITU-R Rec. P.910, the following conditions are recommended: Number of Participants $\geq$ 4 and $\leq$ 40; Viewing Distance: 1-8 $\times$ H, where H is the height of the TV/VDU; Compute Mean Opinion Score (MOS) — see Table 1.1; Spatiotemporal Analysis.

In accordance with the recommended conditions specified in the ITU-R Rec. P.910 subjective evaluation procedures, four individuals participated in an exhaustive assessment (AI and RA tests at QP 22, 27, 32 and 37). Note that two additional individuals participated in preliminary subjective evaluations (AI and RA tests at QPs 17 and 22 only). The viewing distance of the participants from the TV/VDU is 1.5m in all evaluations (1.5m $\approx$ 59.1 inch). The height H of the TV/VDU is 15.7 inch and the viewing distance is approximately 4 $\times$ H (see Figure 1.7).

Concerning the two additional subjective evaluations, the objective is to ascertain if the proposed techniques achieve visually lossless coding (MOS = 5) in the AI and RA tests at QPs 17 and 22 only. To this end, the participants confirmed that an MOS = 5 was recorded in these additional subjective evaluations.

1.3 PhD Thesis: Contributions to Knowledge and Thesis Outline

There exists a research gap in the literature for novel visually lossless coding algorithms, particularly for application in HEVC and for utilisation on high bit depth and HD YCbCr 4:4:4 data. The techniques proposed in this thesis constitute novel approaches as regards the perceptual compression of HD YCbCr 4:4:4, 4:2:2 and 4:2:0 raw video data of various bit depths. In addition to this, the combined objective and subjective evaluations may also be considered a contribution to knowledge. For instance, perceptual quantisation techniques have not been evaluated on different chroma sampling versions and different bit depth versions of the same sequence (e.g., the YCbCr 4:4:4 10-bit version and the YCbCr 4:2:0 8-bit version of DuckAndLegs).

1.3.1 TCPQ (Chapter 3)

In Chapter 3, a novel Transform Coefficient-level Perceptual Quantisation (TCPQ) technique is proposed [21]. TCPQ employs a novel distance parameter in the DCT [32] frequency domain. TCPQ exploits the well-established MTF characteristics of the HVS [1] without the need to develop a complex HVS-based perceptual model. The functionality of TCPQ is somewhat similar to the workings of HVS-related Quantisation Matrices (QMs) [33, 34]; i.e., quantising a transform coefficient according to its frequency content. At its core, TCPQ takes into account the distance of an AC transform coefficient from the DC coefficient in luma and chroma Transform Blocks (TBs). Therefore, TCPQ quantises more coarsely the least perceptually relevant transform coefficients (i.e., the high frequency AC coefficients). Conversely, TCPQ preserves the integrity of the DC coefficient and the very low frequency AC coefficients. Compared with RDOQ [35, 36], which is the most widely used transform coefficient-level quantisation technique in video coding, TCPQ successfully achieves visually lossless coding at low bitrates.

1.3.2 CCCPQ and CBPQ (Chapter 4)

Psychophysical experiments reveal that the HVS is not significantly sensitive to distortions in high variance (busy) regions of picture data, such as quantisation induced noise [37, 38]; this is known as (HVS-related) spatial masking. Therefore, in accordance with this spatial masking component of the HVS, we exploit this phenomenon in the novel techniques proposed in Chapter 4. To this end, two perceptual quantisation techniques are proposed for HEVC at the CU level and at the CB level. Cross-Colour Channel Perceptual Quantisation (CCCPQ) [22] and Coding Block-level Perceptual Quantisation (CBPQ) [23], respectively, are the names of these techniques. CCCPQ and CBPQ improve upon on the luma-only perceptual quantisation technique in HEVC, named AdaptiveQP [39]-[41]. That is, we extend AdaptiveQP to facilitate the perceptual quantisation of data in the chroma channels.

CCCPQ perceptually adjusts the QP at the CU level by computing the spatial variances across the Y CB, the Cb CB and the Cr CB. Therefore, this mechanism equates to cross-colour channel dependency for QP selection at the CU level. With CCCPQ, the CU-level QP is perceptually increased if high spatial variance is detected (across all colour components in a CU). In other words, the QP is adjusted at the CU level and, therefore, the corresponding QP is applied to all three CBs in the CU (i.e., one QP per CU is signalled to the decoder).

Designed primarily for high bit depth 4:4:4 data, CBPQ improves upon CCCPQ by separately adjusting the QP of each CB according to the spatial variances in the Y CB, the Cb CB and the Cr CB (i.e., three QPs per CU are signalled to the decoder). Furthermore, separately adjusting the QP of each CB in a CU is arguably a more intuitive method for perceptually quantising high bit depth 4:4:4 video data. CBPQ exploits the CU-level chroma QP offset and signalling mechanism standardised by JCT-VC [18, 19]; this facilitates an efficient encoder side implementation of the proposed CBPQ technique. Compared with AdaptiveQP, both CCCPQ and CBPQ achieve visually lossless coding at low bitrates. Note, however, that CBPQ attains higher coding efficiency than CCCPQ (also, CBPQ is tested specifically on high bit depth YCbCr 4:4:4 video data).

1.3.3 Pixel-PAQ (Chapter 5)

Inspired by the research gap within the area of JND-based visually lossless coding in HEVC, we propose a high bit depth capable, JND-based perceptual quantisation technique (for both luma and chroma data) in Chapter 5; the technique is named Pixel Perceptually Adaptive Quantisation (Pixel-PAQ). The proposed technique improves upon a luma-only JND-based perceptual quantisation method developed for HEVC by Naccari and Mrak; Intensity Dependent Spatial Quantisation (IDSQ) is the name of this technique [42].

Pixel-PAQ exploits JND-based luminance adaptation and chrominance adaptation, the objective of which is to accomplish visually lossless coding at extremely low bitrates. Based on a novel JND model, Pixel-PAQ achieves its objective by perceptually increasing the Y QStep, the Cb QStep and the Cr QStep at the CB level. This is accomplished by employing parabolic piecewise functions to facilitate luminance masking and chrominance masking (these functions are utilised to perceptually adjust the aforementioned CB-level QSteps). The luminance adaptation and chrominance adaptation psychophysical experiments conducted by X. H. Zhang et al. [43] and G. Wang et al. [44], respectively, form the basis of the proposed piecewise functions.

In addition, Pixel-PAQ accounts for the bit depth of the raw YCbCr video data, which constitutes a novel extension to Naccari's and Mrak's IDSQ technique in [42] (i.e., IDSQ accounts for 8-bit luma data only). Pixel-PAQ is compatible with raw YCbCr video data of any bit depth. This is significant because YCbCr 4:4:4 video data, for example, usually contains higher bit depth samples — i.e., 10-bits per sample per channel — in comparison with chroma subsampled 4:2:0 video data. Furthermore, as is the case with the proposed CBPQ technique in Chapter 4, the CU-level chroma QP offset and signalling mechanism is also exploited by Pixel-PAQ.

Compared with IDSQ, Pixel-PAQ attains bitrate reductions of up to 72.7% over five initial QP data points. This includes a 75% bitrate reduction, which was achieved in the RA QP = 22 test. Moreover, Pixel-PAQ attains visually lossless coding at extremely low bitrates.

1.3.4 List of Publications Relevant to the Thesis

L. Prangnell, "Visually Lossless Coding in HEVC: A High Bit Depth and 4:4:4 Capable JND-Based Perceptual Quantisation Technique for HEVC," *Elsevier Signal Processing: Image Communication Journal*, vol. 63, pp. 125-140, 2018. (Chapter 5).

L. Prangnell and V. Sanchez, "JND-Based Perceptual Video Coding for 4:4:4 Screen Content Data in HEVC," in *Proceedings of the 43rd IEEE International Conference on Acoustics, Speech and Signal Processing*, Calgary, Alberta, Canada, 2018. In Press. (Chapter 5).

L. Prangnell, M. Hernández-Cabronero and V. Sanchez, "Coding Block Level Perceptual Video Coding for 4:4:4 Data in HEVC," *in Proceedings of the 24th IEEE International Conference Image Processing,* Beijing, China, 2017, pp. 2488-2492. (Chapter 4).

L. Prangnell, M. Hernández-Cabronero and V. Sanchez, "Cross-Color Channel Perceptually Adaptive Quantization for HEVC," in *Proceedings of the 27th IEEE Data Compression Conference*, Snowbird, Utah, USA, 2017, pp. 456. (Chapter 4).

L. Prangnell, V. Sanchez and R. Vanam, "Adaptive Quantization by Soft Thresholding in HEVC," in *Proceedings of the 31st IEEE Picture Coding Symposium*, Cairns, Queensland, Australia, 2015, pp. 35-39. (Chapter 3).

Chapter 2. Literature Review

2.1 Related Background Information

2.1.1 The Physics of Visible Light

Every form of electromagnetic radiation in the electromagnetic spectrum, including visible light, microwaves and gamma rays, is a different manifestation of light; the photon is the fundamental particle of light. As is the case with all elementary particles, the photon behaves simultaneously as both a particle and a wave; (i.e, the wave-particle duality phenomenon in nature, as discovered in the field of quantum mechanics [45]). Visible light is a small range of light in the electromagnetic spectrum that is visible to human observers by virtue of the HVS. The established photon wavelength range in the visible light spectrum is approximately 380 to 750 nm, which equates to a frequency range of 668 to 484 THz, respectively [45]. Note that the photon energy of visible light ranges from 2 to 2.75 eV [45]; therefore, the energy of a photon is inversely proportional to its wavelength.

Colour vision in humans equates to the combined visual perception of the different luminance levels and photon energies emitted from either natural or synthetic visible light sources. Every aspect of colour that humans visually perceive is ultimately contingent upon the natural processes of visible light and the subsequent biological processing of such. In other words, colour is the subjective interpretation of electromagnetic radiation in the spectrum of visible light [46]-[48].

As previously mentioned, the photon acts as both a wave and a particle. Photon energy E is measured in either J or eV [25]; E is quantified in (2.1):

$$E = \frac{h \cdot c}{\lambda} \tag{2.1}$$

Table 2.1: The wavelength, frequency and energy of photons in the visible light range of the electromagnetic spectrum. This range of photon energies (and the corresponding wavelengths) manifests perceptually as a range of colours in the visual systems of humans, African monkeys, apes and chimpanzees.

Colour Perception	Wavelength λ	Frequency v	Energy E
Violet	380-450 nm	668-789 THz	2.75-3.26 eV
Blue	450-495 nm	606-668 THz	2.50-2.75 eV
Green	495-570 nm	526-606 THz	2.17-2.50 eV
Yellow	570-590 nm	508-526 THz	2.10-2.17 eV
Orange	590-620 nm	484-508 THz	2.00-2.10 eV
Red	620-750 nm	400-484 THz	1.65-2.00 eV

where h and c are physical constants. Planck's constant $h = 6.626\times10^{-18}$ J seconds, which computes the quantum of action. Constant $c = 3\times10^{8}$ m/s and approximates the speed of light in a vacuum. Lambda λ refers to the wavelength of a photon in nm. Note that 1 J = 6.242×10^{18} eV.

Utilising the formula in (2.1), we can quantify the photon energy of electromagnetic radiation — visible light range — that humans perceive as red (red light), denoted as E_{red}, in (2.2).

$$E_{\text{red}} = \frac{\left(6\times10^{-34}\,\text{J}\cdot\text{s}\right)\times\left(3\times10^{8}\,\text{m/s}\right)}{700\times10^{-9}\,\text{m}} = 2.8\times10^{-19}\,\text{J} = 1.8\text{eV} \tag{2.2}$$

As shown in (2.2), assuming the wavelength of 700 nm, the photon energy of red light is approximately 1.8 eV (see Table 2.1). The number of photons N emitted per second by a visible light source is contingent upon the energy of the visible light source P, measured in J, and also the energy of the photon [49, 45], which is given by (2.3).

$$N = \frac{P}{E} \tag{2.3}$$

2.1.2 The Human Visual System

From the perspective of the Darwinian paradigm of evolutionary biology, the HVS is the product of billions of years of evolution by natural selection [50]. The interaction of photons with the retinal photoreceptor systems facilitates colour vision, wherby the photons are biologically converted into electrical signals in the retina [51, 52]. Visual perception of colour depends on the level of excitation of the different cones. Furthermore, in general terms, the visual cortex system in the brain is responsible for differentiating the signal response received from the Long, Medium and Short (L, M, S) cones. This facilitates the discernment of a vast range of signals that are perceived in the form of a wide range of colours.

With a focused concentration on retinal photoreceptors, as shared by all species in the taxonomic order of primate (including humans), rods and cones constitute the key photoreceptors. The retinal photoreceptor system is dominated by rods (120,000,000 units) compared with cones (6,400,000 units). Rods are specialised for low visible light conditions [51, 52]. When subjected to higher intensities of visible light the transmitter release stops because the rod's response to the visible light is much slower than the cone's response. Cones are the retinal photoreceptors that facilitate colour vision and colour perception. They are able to adapt to a vast variety of visible light intensities [51, 52].

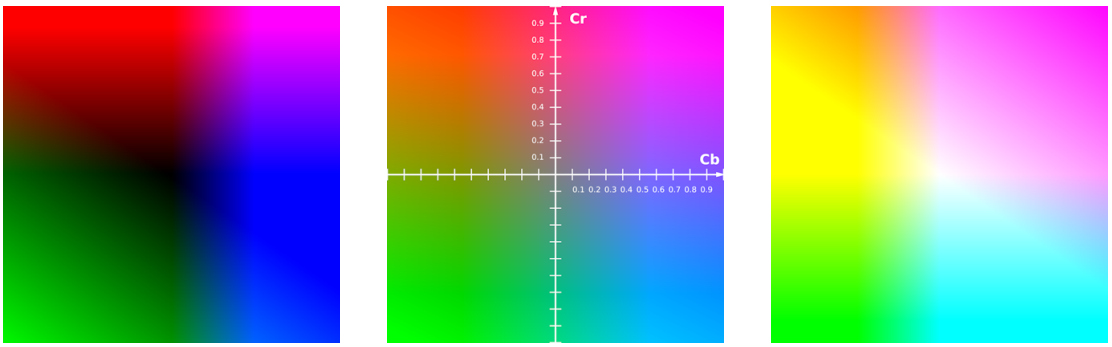
In terms of the population of cones, empirical experiments have revealed that 64% are sensitive to photons perceived by the HVS as red, 32% green and 4% blue (i.e., trichromatic colour vision). There are three classifications of retinal cone: L, M, S, each of which contains the transmembrane protein opsin and the molecule chromophore, which are the constituents of photosensitive visual pigments. These pigments are especially sensitive to photons within the following approximate photon wavelength ranges: 650 nm (L), 510 nm (M) and 475 nm (S), which humans interpret as red, green and blue, respectively [50]-[52]. In essence, the relationship between visible light and the HVS catalysed the emergence of the Red, Green, Blue (RGB) colour model and the corresponding YCbCr colour space.

2.1.3 RGB and YCbCr

The Young-Helmholtz theory of trichromatic colour vision is the scientific basis for the RGB colour space. The RGB colour model is an additive tristimulus colour model that is ubiquitous in computer science applications and consumer electronics devices. It amalgamates colour from the following primary colours: red, green and blue, which results in a range of colours depending on the corresponding sample intensity, colour gamut and the associated bit depth. In terms of the physical (hardware) pixels built TVs and monitors, each pixel in these devices contain visible light sources. Therefore, the physics of visible light and the biology of colour vision always apply.

The interaction of the HVS with luminance and photon energy, emitted from the VDU or TV (light source), is the physical process by which visible light and the associated luminance is visually perceived as brightness and colourfulness. RGB-based colour values can be represented by normalised arithmetic, percentage or base-10 integer representations of binary numbers. Expressed as a triplet, the binary representations of R, G and B data are dependent on the bit depth of each channel. There are 2^b sample intensities in each colour channel, where b denotes the bit depth of the data.

For a bit depth of 8-bits per sample per channel (i.e., 24-bits per sample), the integer value ranges are as follows: $R \in [0,255]$, $G \in [0,255]$ and $B \in [0,255]$. In this example, R=0, G=0 and B=0 represents absolute black (low energy). Conversely, R=255, G=255 and B=255 represents absolute white (high energy). For image or video data with higher bit depths (e.g., 48-bits per sample), this equates to a greater number of colours in each sample. For 48-bit image or video data, the value ranges are as follows: $R \in [0,65535]$, $G \in [0,65535]$ and $B \in [0,65535]$, where R=65535, G=65535 and B=65535 represents absolute white. The YCbCr colour space is a colour transformation from a given RGB colour space; YCbCr comprises one luma channel (Y) and two chroma channels (Cb and Cr). Note that YCbCr is often confused with the YUV colour space; they are closely related, however. YCbCr is designed for digital imaging and video, whereas YUV is designed for analogue imaging and video; YCbCr is a scaled version of YUV.



(a) (b) (c)

Figure 2.1: The chrominance *Cb′* and *Cr′* colour planes using the RGB colour gamut range; in these examples, the *Cb′* and *Cr′* planes are in the range [−1,1] for a normalised *Y′* value (i.e., $Y' \in [0,1]$). Subfigure (a) shows the Cb and Cr colour planes when $Y' = 0$; subfigure (b) shows the Cb and Cr colour planes when $Y' = 0.5$; subfigure (c) shows the Cb and Cr colour planes when $Y' = 1$.

In YCbCr, luma (Y) refers to an achromatic colour channel that is derived via an approximation of gamma-corrected luminance. The human perception of the brightness of colour in the luma channel is conceptualised as relative luminance, in which the values are normalised to 1 or 100. On a percentage scale, 0% represents absolute black and 100% represents absolute white; moreover, the luma channel contains the vast majority of the finer detail in an image. Chrominance (Cb and Cr) refers to the difference colour channels, which are as follows: blue difference (Cb) and red difference (Cr) with reference to the luma (Y) channel; Cb and Cr collectively correspond to the saturation of the colour in an image (see Figure 2.1).

In a mathematical sense, *Y*, in YCbCr, which is referred to as luma, corresponds to the weighted sum of RGB values (not gamma corrected). The gamma corrected version is denoted as *Y′*. Concentrating on the gamma corrected version, *Y′*, from ITU-R BT.709 [53], is computed in (2.4).

$$Y' = (0.2627 \cdot R') + (0.6780 \cdot G') + (0.0593 \cdot B') \tag{2.4}$$

The parameter values applied to the gamma corrected RGB data (*R′*, *G′* and *B′*) are weights determined by a luminosity function. It has been shown, in empirical testing, that humans are perceptually more sensitive to green in terms of brightness perception, in which case *G′* is assigned the largest weight. Conversely, *B′* is assigned the lowest weight.

Cb' and Cr' are colour difference channels with reference to the Y' colour channel; the perceived hue and saturation in the Cb' and Cr' is dependent on the value of the Y' channel (see Figure 2.1). The gamma corrected versions of Cb and Cr, from ITU-R BT.709 [53], are computed in (2.5) and (2.6), respectively.

$$Cb' = \frac{B - Y'}{1.8814} \tag{2.5}$$

$$Cr' = \frac{R - Y'}{1.4746} \tag{2.6}$$

Note that YCbCr 4:4:4, 4:2:2 and 4:2:0 refer to the sampling (resolution) ratios of the chrominance data with respect to the resolution of the luma data. In 4:4:4 data, the Cb and Cr data is the same resolution as the Y data (i.e., no chroma subsampling). The YCbCr 4:2:2 and 4:2:0 versions use spatial chroma subsampling, which is a form of compression. In 4:2:2 data, the Cb and Cr data is half the resolution of the Y data, and in 4:2:0 data, the Cb and Cr data is quarter the resolution of the Y data.

2.2 Visually Lossless Coding and Review of Related Work

Perceptual considerations in video compression have always been a focus of concern during the official standardisation of video coding platforms; however, HVS-related psychovisual redundancy reduction techniques (e.g., JND-based luminance masking) are neglected, thus leaving significant room for improvement. H.263 [54], Moving Picture Experts Group version 4 (MPEG 4) [55], AVC [7, 8] and HEVC all include RDT-based coding techniques including intra prediction, inter prediction, residual transform coding and scalar quantisation; all of these techniques reduce perceptual redundancies that are present in the raw video data. Inter prediction reduces spatiotemporal redundancies between frames, transform coding preserves the most perceptually significant transform coefficients and scalar quantisation discards the least perceptually important coefficients (albeit to a limited degree). In contemporary perceptual video and image compression research, increasingly sophisticated visually lossless coding techniques are being proposed — e.g., luminance spatial masking methods based on a JND model — to maximise bitrate reductions [56, 57, 1, 42].

In early video coding research, scientists discovered that spatially subsampling the analogue chrominance components (U and V) in YUV analogue video data is usually not noticeable to the HVS [58] (note that this was discovered on obsolete CRT-based visual display technologies). This gave rise to YUV 4:2:0 (analogue) and also YCbCr 4:2:0 (digital) chroma subsampling. Concentrating on the coding of raw digital data, this means that YCbCr 4:2:0 data, for example, has already been compressed prior to being further compressed by a video coding platform such as HEVC and/or AVC. It is important to note that, although the aforementioned coding techniques, including GOP-based inter prediction, transform coding and scalar quantisation are designed to reduce certain redundancies in raw video data, these techniques can be perceptually optimised in order to maximally reduce perceptual redundancies in the compressed video data. It is important to affirm the fact that subsampling compression artifacts in 4:2:0 data are typically conspicuous on contemporary 4:4:4 and HDR capable TVs and VDUs. This is the reason why 4:4:4 data is becoming increasingly popular.

Focusing on scalar quantisation in HEVC, recall that the default quantisation technique in HEVC is URQ [5, 13, 17]. Because URQ equally quantises entire TBs of luma and chroma transform coefficients, it does not intrinsically possess the capacity to consider the perceptual importance of individual coefficients in a TB. This has engendered the emergence of coefficient-level quantisation methods including RDOQ [35, 36] and perceptual quantisation schemes based on a JND model.

In the JCT-VC HEVC HM reference codec [29, 30], the video coding algorithms in HEVC HM are based primarily on RDT. Note that RDT is a mathematical model used in lossy coding methods (that are not perceptually optimised) to find a balance between bitrate and distortion without the need to employ subjective evaluations. Consequently, objective quality measurements in HEVC lossy video coding applications are based on MSE [59]; i.e, the MSE of the reconstructed samples compared with the samples in the raw data. During the standardisation of HEVC, the spatially orientated PSNR visual quality metric — which is a logarithmic metric based on MSE — is employed to quantify the reconstruction quality of compressed video data. However, the contemporary perceptual coding literature confirms that PSNR has a very poor correlation with human visual perception [1]. It can be argued that this is because MSE merely computes the average of the squares of the deviations [59].

Studies have shown that a compressed video with a PSNR measure of 40 Decibels (dB), or above, typically constitutes visually lossless coding. That is, a coded video with a PSNR $\geq$ 40 dB is perceptually indistinguishable from the raw video data. Moreover, using the example of PSNR $\geq$ 40 dB for visually lossless coding, this also implies that targeting a reconstruction quality of PSNR > 40 dB (i.e., PSNR = 50 dB) is superfluous. That is, unnecessary bits are wasted by achieving the superior mathematical reconstruction quality required for the PSNR = 50 dB measurement. This is the main reason why subjective evaluations are extremely important.

In addition to the primary objective of improving coding efficiency, most lossy video coding algorithms employed in HEVC HM are developed with an emphasis on increasing the PSNR values in the compressed video data. These algorithms include RDO [60, 61], RDOQ [35, 36], Deblocking Filter (DF) [62] and Sample Adaptive Offset (SAO) [63]. RDO, RDOQ, DF and SAO are effective methods in terms of increasing PSNR values for the reconstructed video. However, the reconstruction quality improvement attained by these techniques, as quantified by PSNR, is usually negligible in terms of how the human observer interprets the perceived quality of the compressed video data.

Subjective evaluations prioritise, above all else, the human observer with respect to assessing the reconstruction quality of a compressed video sequence. In essence, the human observer is the ultimate judge of the visual quality of a compressed video sequence. Therefore, human subjective quality evaluations are critically important in terms of assessing the perceptual quality of compressed video data. As previously implied, the main objective with visually lossless coding methods is to reduce the number of bits required to store each sample without incurring a visually perceptible loss of quality in the compressed data. Note that with visually lossless coding techniques (either JND-based or otherwise), PSNR measurements are not considered to be important in terms of quantifying the perceptual quality of a reconstructed sequence. In such cases, the PSNR metric is typically utilised for quantifying the degree to which PSNR values can be decreased before the associated compression artifacts in the coded video are discernible by the human observer.

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

(a)

16	16	16	16	17	18	21	24
16	16	16	16	17	19	22	25
16	16	17	18	20	22	25	29
16	16	18	21	24	27	31	36
17	17	20	24	30	35	41	47
18	19	22	27	35	44	54	65
21	22	25	31	41	54	70	88
24	25	29	36	47	65	88	115

(b)

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

(c)

Figure 2.3: QMs utilised in the JPEG standard. Subfigures (a), (b) and (c) show a flat QM, the luma QM and the chrominance QMs, respectively.

The built-in visual distortion masking properties are exploited in visually lossless coding techniques; this includes the aforementioned spatiotemporal visual masking phenomenon [1, 64]. By default, video coding standards, including HEVC, already employ redundancy reduction methods including spatiotemporal prediction, transform coding and scalar quantisation; however, these techniques are not perceptually optimised by default. Visually lossless coding methods are coveted because, by definition, they often dramatically reduce the number of bits required to store a sample (e.g., a 10-bit sample in the luma channel of a YCbCr raw data file).

A primitive example of visually lossless coding at the transform coefficient-level is the QM (see Figure 2.3). In the Joint Photographic Experts Group (JPEG) still image coding standard [65], QMs are employed based on a CSF-based psychovisual model. As previously mentioned, the DCT basis functions operate according to the MTF characteristics of the HVS. After decorrelating an image into the frequency domain, the DC transform coefficient and the low frequency AC transform coefficients contain almost all of the important detail of the image; therefore, the very high frequency AC coefficients in luminance data can be discarded (zeroed out). Therefore, the default QMs for the JPEG standard are designed on this principle. In addition to this, the HVS is much less sensitive to compression artifacts in chrominance data. Therefore, the luminance QM differs considerably from the chrominance QM. The luminance QM discards the very high frequency coefficients. Conversely, the chrominance QM discards the medium frequency and all high frequency AC coefficients (see Figure 2.3). This is because it is significantly more difficult for the HVS to detect quantisation-induced compression artifacts in compressed chrominance data.

Wang et al. in [33] developed a HVS-based QM for the JPEG standard; this approach is based on Daly's 2D spatial CSF method [66]. In [33], the authors propose a CSF based QM technique in which Frequency Weighting Matrices (FWMs) are derived. After scaling the FWM values to integer values, the integer QM values are derived. The QM technique by Wang et al. in [33] is employed as the default QM technique in the HEVC standard [34]. A significant issue with this, however, is that this QM method is applied to both luma and chroma coefficients during the frequency dependent scaling process in HEVC, which is not appropriate. Similarly, in [67], Naccari and Pereira propose a CSF-based QM technique for the AVC standard. In this work, the luminance JND visibility threshold is applied to both luma and chroma data. This is based on the assumption that the luminance and chrominance spatial CSFs exhibit similar properties; it is well established, however, that the HVS is vastly more sensitive to luminance (achromatic) contrast.

RDOQ [35, 36] is a state-of-the-art example of a transform coefficient-level quantisation technique. The core objective with RDOQ technique is to establish an optimal quantisation level for each transform coefficient in luma and chroma TBs. RDOQ quantifies the quantisation-induced distortion and also the number of bits required to encode the corresponding quantised transform coefficient. Based on these two values, the RDOQ chooses an optimal coefficient value, which is determined by finding an appropriate trade-off between bitrate and quantisation-induced distortion (i.e., the RD cost). RDOQ is widely adopted and it is the default quantisation technique utilised in HEVC (in combination with URQ). Due to the wide adoption of RDOQ, QMs are seldom utilised in contemporary video coding applications. It is worth noting that very little research is conducted on QMs in present research; moreover, they are disabled by default in the HEVC HM reference software; QMs excluded from all JCT-VC HEVC common test conditions. Although RDOQ is an advanced coefficient-level quantisation technique, it has been designed with a significant emphasis on improving the mathematical reconstruction quality of the coded data as quantified by PSNR (compared with URQ and QMs, for example). RDOQ is not perceptually optimised; it does not take into account HVS-related psychovisual redundancies.

Numerous psychophysical experiments have revealed that the HVS is less sensitive to quantisation-related distortions within regions of image data that comprise high spatial variations [68, 37, 1]; this constitutes a form of visual masking in the spatial domain. Therefore, in comparison with standard uniform quantisers including URQ, perceptual quantisers of this nature can be designed to exploit the aforementioned spatial domain visual masking phenomenon of the HVS. In the context of video coding and perceptual quantisation, this means that higher levels of quantisation can be applied to high spatial activity regions in the frames of a sequence, thus potentially giving rise to bitrate reductions.

In HEVC, AdaptiveQP [39]-[41] is an example of a HVS-based perceptual quantisation method which can exploit the spatial masking phenomenon of the HVS. A notable shortcoming of AdaptiveQP, however, is that it adjusts the QP of an entire $2N \times 2N$ CU based solely on the sample variance in the constituent luma CB. If, for example, AdaptiveQP is employed to decrease overall bitrates by increasing the CU level QP (based only on the sample variance in a luma CB), an inappropriate QP adjustment may be applied to the chroma Cb and Cr CBs; this is particularly pertinent to high bit depth 4:4:4 video data. The chroma Cb and Cr channels in, for example, 10-bit 4:4:4 video sequences may contain higher sample variances compared with the sample variances in the luma channel. This is by virtue of the nature of 10-bit 4:4:4 video data; i.e., 30-bits per sample in addition to an absence of chroma subsampling.

Visually lossless coding based on a JND model is presently of considerable interest in video coding and image coding research. JND refers to the maximum visibility threshold before lossy compression distortions become perceptually conspicuous to the human observer [1]. In JND-based perceptual coding techniques, a JND visibility threshold is typically established. Therefore, assuming that the threshold is not exceeded, visually lossless coding is successfully achieved; this is generally confirmed with the use of subjective evaluations (e.g. MOS = 5). The vast majority of JND techniques in video compression applications target the spatiotemporal domain, the frequency domain or a combination of the two. In HEVC, JND-based perceptual quantisation (for both luma and chroma data), for example, can considerably reduce HVS-related psychovisual redundancies that are present in each channel within raw YCbCr video data.

Mannos' and Sakrison's pioneering work in [69] formed a useful foundation for all frequency domain luminance CSF-based JND techniques which target HVS-based redundancies in luminance image data. In [70], Ahuma and Peterson devise the first DCT-based JND technique, in which a luminance spatial CSF is incorporated. In [71], Watson expands on Ahuma's and Peterson's work by incorporating luminance masking and contrast masking into the DCT-based method; note that power functions corresponding to Weber's law are utilised in this method.

Chou and Chen propose a pioneering spatial domain JND profile in [72], in which luminance masking and contrast masking functions are proposed for utilisation in the spatial domain (luma samples of 8-bit precision). This method is based on average background luminance and luminance adaptation. The authors further expand on this method in [73] by adding a temporal masking component, in which inter-frame luminance is exploited. X. Yang et al. in [74] propose a spatiotemporal domain JND contribution to eradicate the overlapping effect between luminance masking and contrast masking effects. This technique also includes a filter for motion-compensated residuals, in which they employ a modified version of Chou's and Chen's spatiotemporal domain JND methods.

X. H. Zhang et al. in [43] propose a DCT-based JND technique, in which a luminance adaptation parabolic piecewise function is derived. It is established that JND-based compression can be accomplished by employing luminance masking mechanisms for light and dark regions in luminance data. In [75], Jia et al. propose a DCT-based JND technique founded upon a CSF-based temporal masking effect. Wei and Ngan in [76] introduce a DCT-based JND contribution for video coding, in which the authors incorporate luminance masking, contrast masking and temporal masking effects into the technique. The luminance masking component is modelled as a piecewise linear function. The contrast masking aspect is contextualised as edge and texture masking; the temporal masking component quantifies temporal frequency by taking into account motion direction. Chen and Guillemot in [77] propose a spatial domain foveated masking JND technique, which is the first time that image fixation points are taken into account in JND modelling.

More recently (i.e., from 2013 onwards), several JND-based perceptual coding techniques have been proposed, mainly for the HEVC standard. In [42], Naccari and Mrak propose a luminance JND-based perceptual quantisation method for HEVC. It is designed to exploit luminance adaptation-related intensity masking and is based on the research conducted by X. H. Zhang et al. in [43]. Y. Zhang et al. in [78] expand on Naccari's and Mrak's JND technique in [42] by adopting it for HDR-related tone mapping applications in HEVC.

Kim et al. in [79] propose a hybrid frequency domain and spatiotemporal domain JND technique for HEVC. This method combines the following visual masking properties to create a JND visibility threshold: spatial CSF, luminance masking, luminance temporal masking and contrast masking. Note that the spatial domain aspect of the technique is utilised for transform skip mode in HEVC. In [80], Bae et al. propose a JND-based perceptual video coding scheme for HEVC, in which the JND visibility threshold adapts to the size of the transform. In addition, the frequency sub-band of luma transform coefficients is taken into account. As with previously proposed JND techniques, the JND profile in this method is based on luminance spatial CSF, luminance masking and contrast masking.

Wu et al. in [81] propose a luminance-based JND model for images with pattern complexity. This technique exploits the fact that the HVS is less sensitive to compression artifacts in high variance regions of luminance data; therefore, these regions can be compressed to a much higher degree than low variance regions. This method is based primarily on luminance contrast and pattern masking.

In [82], Bae et al. propose a DCT-based JND technique which amalgamates luminance-based temporal masking and foveated masking. These visual masking effects are combined with spatial CSF, luminance spatial masking and contrast masking to create a full JND profile. G. Wang et al. in [44] propose a multi-channel DCT-based JND technique for HEVC; this combines spatial CSF, spatial masking, contrast masking and temporal masking for all colour channels in YCbCr video data.

As is evident in the vast majority of the previously proposed perceptual coding techniques reviewed in this section, the JND of chrominance data is typically neglected; it is important to note, however, that the method proposed by G. Wang et al., in [44], does account for the chrominance data. Moreover, many of these methods share one or more of the same features, which are as follows: luminance-based spatial masking, luminance-based contrast masking, luminance-based temporal masking and luminance-based spatial CSF. In many of the JND techniques reviewed in this section, the JND visibility threshold for chrominance data is treated as identical to the JND visibility threshold for luminance data. This is a significant shortcoming of the previously proposed techniques because chrominance data is demonstrably different from luminance data. Chrominance data can be compressed to a significantly higher degree than luminance data; therefore, this fact should be exploited in visually lossless coding techniques.

In addition to the aforementioned issue concerning the absence of accounting for chrominance JND, other issues exist that are not considered in previously proposed JND techniques. For example, Yang et al., Chen and Guillemot, Kim et al. and Wu et al., in [74], [77], [79] and [81] respectively, adopt the luminance-based spatiotemporal visual masking functions derived by Chou's and Chen's techniques in [72, 73]. The issue here is as follows: the psychophysical experiments undertaken by Chou and Chen were conducted in 1995-96 on obsolete visual display technologies (i.e., an SD and low resolution 19 inch CRT monitor). More specifically, Chou and Chen in [72] derived the spatial domain average background luminance function based on psychophysical experiments conducted on a very old and obsolete 19 inch, SD and low resolution CRT monitor. Therefore, Chou's and Chen's corresponding luminance masking and contrast masking functions may require further investigation. This is because the derived JND visibility thresholds may prove to be significantly different if the corresponding subjective evaluations were to be performed on contemporary visual display technologies (e.g., a state-of-the-art monitor which supports HD, Ultra HD, HDR, WCG and 4:4:4 video data).

Another issue with previously proposed JND methods — with the exception of Y. Zhang's High Dynamic Range (HDR)-related tone-mapping extension [78] of Naccari's and Mrak's IDSQ technique in [42] — is the fact that they are designed for raw 24-bit YCbCr video data (i.e., 8-bits per channel data). This equates to the fact that most of the aforementioned empirical parameters in the luminance masking, contrast masking and temporal masking functions are designed to work with 8-bit precision data only. This is a significant issue considering the fact that high bit depth data (i.e., up to 16-bits per channel data) is becoming increasingly popular. At present, 8-bits per channel data remains ubiquitous; however, this is likely to change in the near future.

Chapter 3. Transform Coefficient-Level Perceptual Quantisation

In this chapter, a transform coefficient-level perceptual quantisation technique is proposed (named TCPQ [21]). TCPQ individually quantises transform coefficients according to a Euclidean distance parameter. It quantises coefficients based on the distance of the individual AC coefficient from the DC coefficient in luma and chroma TBs. TCPQ is thus designed to quantise high frequency AC transform coefficients to a much greater level than the DC coefficient and the low frequency AC coefficients. Moreover, the proposed method is compatible with raw video data of any bit depth and any chroma sampling ratio. TCPQ facilitates a reduction of non-zero quantised coefficients; therefore, the corresponding data can be compressed more efficiently during the entropy coding process with CABAC [20]. The objective and subjective evaluations reveal that TCPQ achieves important bitrate reductions and visually lossless coding.

The rest of the chapter is organised as follows. Section 3.1 includes background information relevant to the proposed method. In section 3.2, the proposed technique is presented. In section 3.3, thorough subjective and objective experimental evaluation results are presented and discussed. Finally, section 3.4 concludes this chapter.

3.1 Related Background

After intra prediction and inter prediction, a finite precision integer approximation of DCT [32] is applied to the corresponding residual signals, from which transform coefficients are derived [12, 13]. The integer approximation of DCT is applied to 8×8 to 32×32 intra residual luma and chroma residual blocks. For inter-predicted residuals, the corresponding integer approximation of DCT is utilised on 4×4 to 32×32 luma and chroma residual blocks. For 4×4 intra residue, an integer approximation of DST is utilised instead of DCT [12, 13]. Recall that the integer DCT and DST schemes in HEVC exploit the MTF characteristics of the HVS. They compact the energy of luma and chroma prediction residual samples into the DC coefficient and the very low frequency AC coefficients [12, 13].

To reiterate, it is well known that the DC coefficient and the low frequency AC transform coefficients are more important than the high frequency AC coefficients (in terms of reconstructing the compressed video signal). Because each coefficient frequency sub-band in a TB constitutes a different level of perceptual importance in a compressed video signal [83], the distance of AC coefficients from the DC coefficient can be quantified in terms of Euclidean distance. That is, the DC coefficient is the starting point and the distance of each AC coefficient from the DC coefficient represents the perceptual importance of the current AC coefficient.

Recall that the URQ technique [5, 13, 17] is the compulsory uniform quantisation method in HEVC. Assuming that chroma QP offsets are not employed in HEVC, the QStep computation for luma data is identical to the QStep computations applied to chroma Cb and Cr data. As such, the luma and chroma QSteps in HEVC, denoted as $QStep_Y$, $QStep_{Cb}$ and $QStep_{Cr}$ are defined in (3.1) to (3.3), respectively:

$$QStep_Y\left(QP_Y\right) = 2^{\frac{QP_Y-4}{6}} \tag{3.1}$$

$$QStep_{Cb}\left(QP_{Cb}\right) = 2^{\frac{QP_{Cb}-4}{6}} \tag{3.2}$$

$$QStep_{Cr}\left(QP_{Cr}\right) = 2^{\frac{QP_{Cr}-4}{6}} \tag{3.3}$$

where QP_Y, QP_{Cb} and QP_{Cr} correspond to the integer luma and chroma Cb and Cr QP values, respectively; they are defined in (3.4) to (3.6), respectively.

$$QP_Y\left(QStep_Y\right) = \left\lfloor 6 \times \log_2\left(QStep_Y\right)\right\rfloor + 4 \tag{3.4}$$

$$QP_{Cb}\left(QStep_{Cb}\right) = \left\lfloor 6 \times \log_2\left(QStep_{Cb}\right)\right\rfloor + 4 \tag{3.5}$$

$$QP_{Cr}\left(QStep_{Cr}\right) = \left\lfloor 6 \times \log_2\left(QStep_{Cr}\right)\right\rfloor + 4 \tag{3.6}$$

Table 3.1: The first six values of *QP*, *QStep*, *m* and *s*.

QP	0	1	2	3	4	5
QStep	0.6300	0.7071	0.7937	0.8909	1.0000	1.1225
m	26214	23302	20560	18396	16384	14564
s	40	45	51	57	64	72

In terms of the quantisation of luma and chroma transform coefficients in HEVC and the association of the QP and QStep with the Multiplication Factor (MF) and the Scaling Factor (SF), the quantised transform coefficient within an $N \times N$ TB, denoted as t, is computed in (3.7):

$$t = \frac{C \cdot m + o}{2^{21+\frac{QP}{6}-\log_2 N}} \tag{3.7}$$

where C denotes the transform coefficient, m corresponds to the MF associated with the QP, o refers to the offset corresponding to the error level incurred by quantisation rounding including the level of deadzone and N denotes the N value of an $N \times N$ TB [5, 13, 17]. QStep values are integer approximated in HEVC. The inverse quantised transform coefficient, denoted as C', is computed in (3.8):

$$C' = \frac{t \cdot s \cdot 2^{\frac{QP}{6}}}{2^{\log_2 N - 1}} \tag{3.8}$$

where s is the SF employed for inverse quantization. The URQ method in HEVC is designed such that coefficients in a TB are equally quantised according to the frame level QP; therefore, a single QP value is applied to an entire TB of transform coefficients. MF m and SF s can be computed in (3.9) and (3.10), respectively.

$$m \approx \left[\frac{2^{14}}{QStep} \right] \tag{3.9}$$

$$s = \left[2^6 \times QStep \right] \tag{3.10}$$

Due to the MF and the associated bitwise operations, the values associated with quantisation and inverse quantisation are quantified without the need for divisions and floating point operations. Moreover, as shown in Table 3.1, the MF is inversely proportional to the QP and the QStep. Therefore, any decreases to the MF will induce greater levels of quantisation. One main objective is to ensure that an increment of *QP* (i.e., *QP* + 1) equates to an increase of *QStep* by approximately 12 % [5, 13, 17].

RDOQ [35, 36], which is dependent on URQ, is enabled by default in the JCT-VC HEVC HM software [29, 30]; it is, therefore, the default quantisation technique when following the common test conditions. RDOQ is a soft decision quantisation method which individually quantises coefficients, in both luma and chroma TBs, based on rate-distortion theory. This is achieved by minimising the rate-distortion Lagrangian cost function [84]. RDOQ is designed to search for an optimal set of quantised coefficients in order to establish a suitable trade-off between bitrate and quantised induced distortion; as such, a calculation for each transform coefficient is performed separately. In essence, RDOQ manipulates the quantised transform coefficients according to the final RD performance [35, 36]; therefore, it significantly outperforms URQ in terms of reducing bitrates.

In luma and chroma TBs of size $N \times N$, each transform coefficient C with RDOQ is quantised to three level values, which are as follows: 0, l_1 and l_2. According to [35], for each transform coefficient position in a TB, the Lagrangian cost of each value of C is calculated when the quantisation level value, denoted as l_i, is equal to 0, l_1 or l_2. When C is quantised to value l_i, the Lagrangian cost $J(\lambda, l_i)$ is computed as follows in (3.11):

$$J(\lambda, l_i) = r(C, l_i) + \lambda \cdot b(l_i) \tag{3.11}$$

where λ denotes the Lagrangian multiplier (the value is computed in [35], where r denotes the quantisation error if C is quantised to level l_i and where b corresponds to the number of bits required to code l_i. Variables l_1 and l_2 are computed in (3.12) and (3.13), respectively.

$$l_1 = \left\lfloor |C| \cdot \frac{m}{2^{15+\frac{QP}{6}}} \right\rfloor \tag{3.12}$$

$$l_2 = l_1 + 1 \tag{3.13}$$

Recall that m is the MF, as computed in (3.9). The final quantised level, denoted as q, is computed in (3.14). Therefore, the Lagrangian cost function is updated to $J(\lambda,q)$.

$$q = \arg\min J\left(\lambda, l_i\right) \tag{3.14}$$

A notable drawback of RDOQ is the computational complexity associated with the rate-distortion decisions it carries out. Recall that it is designed primarily to select an optimal quantisation level, with (3.11), to find a suitable trade-off between rate and distortion; this process alone requires significant computational complexity [36].

3.2 Proposed TCPQ Technique

The proposed TCPQ technique is a relatively simple luma and chroma transform coefficient-level perceptual quantisation method. TCPQ is designed to quantise high frequency AC coefficients more aggressively than the corresponding DC coefficient and low frequency AC coefficients. More specifically, TCPQ quantises a transform coefficient individually according to its importance in terms of its contribution to the reconstruction quality of the compressed video data. The level of quantisation applied to an individual AC coefficient is modified according to its position in relation to its Euclidean distance from the DC coefficient. An important feature of TCPQ is the fact that the distance parameter values change according to the size of the transform; therefore, TCPQ adapts to the size of the luma and chroma TBs. In essence, TCPQ has been designed to achieve visually lossless coding at low bitrates. Due to the wide adoption of RDOQ, we compare TCPQ with RDOQ in the experimental evaluations. Moreover, comparing the performance of TCPQ with RDOQ also implies comparing TCPQ with URQ, primarily because RDOQ significantly outperforms URQ alone.

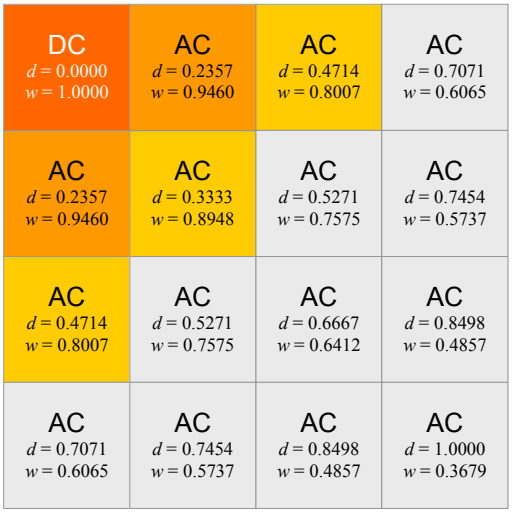


Figure 3.1: A graphical representation of the normalised Euclidean distance d and weight w associated with transform coefficient locations in a 4×4 TB. The darker orange, lighter orange, yellow and grey squares correspond to the DC coefficient, low frequency AC coefficients, medium AC frequency coefficients and high frequency AC coefficients, respectively.

By decreasing the MF, the corresponding larger QStep value is employed primarily for quantising high frequency AC coefficients more aggressively, thus giving rise to perceptual quantisation. More specifically, the level of quantisation applied to each transform coefficient is modified indirectly by weighting — and subsequently decreasing — the corresponding MF value according to a normalised Euclidean distance parameter (see examples in Figure 3.1 and Figure 3.3). The distance parameter is a constituent of an exponentially decreasing function. It is employed for decreasing the MF and subsequently increasing the levels of quantisation applied to each AC coefficient.

The proposed technique reduces non-zero quantised coefficients. Therefore, a significant decrease in bitrates can be accomplished without incurring a perceptually discernible decrease in reconstruction quality; this is confirmed in the objective and subjective evaluations. To reiterate, TCPQ is designed with the simple objective of modifying the MF m, such that the resulting modified MF value, denoted as m', increases the QStep applied to each AC transform coefficient according to its distance from the DC coefficient, thus allowing for an efficient encoder side implementation. Note that, in the proposed TCPQ technique, modified MF m' in (3.15) replaces unmodified MF m in (3.7). The modified MF m' is computed in (3.15):

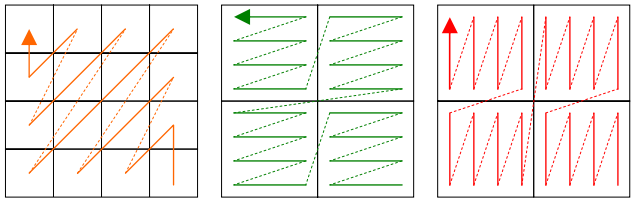


Figure 3.2: A graphical representation of the transform coefficient scan patterns employed in HEVC. Subfigure (a) shows a diagonal reverse scan pattern used to process coefficients within a 4×4 TB. Subfigures (b) and (c) show the horizontal and vertical reverse scan patterns, respectively, for the processing of the constituent 4×4 SBs in 8×8 TBs.

$$m' = m \cdot w \tag{3.15}$$

where w corresponds to an exponentially decreasing function weight. Recall that w modifies the MF for transform coefficients in both luma and chroma TBs; weight w is quantified in (3.16):

$$w(d) = e^{-(d)^2} \in [0,1] \tag{3.16}$$

where d is the normalised Euclidean distance between the position of the current AC transform coefficient in an $N \times N$ TB. Variable d is calculated in (3.17):

$$d = \sqrt{\frac{(x_1 - x_2)^2 + (y_1 - y_2)^2}{(x_1 - x_{\max})^2 + (y_1 - y_{\max})^2}} \in [0,1] \tag{3.17}$$

where (x_1, y_1), (x_2, y_2) and $(x_{\max}, y_{\max})$ represent the (x,y) coordinates of the DC coefficient, the current coefficient and the farthest AC coefficient, respectively. The DC coefficient is at position $x = 0$, $y = 0$ in both luma and chroma TBs.

The proposed technique is suitable for, and can be utilised with, the current scan patterns used for CABAC entropy coding in HEVC (see Figure 3.2). Furthermore, it is important to reiterate the fact that TCPQ is compatible for utilisation with luma and chroma TBs of any size (i.e., from 4×4 samples to 32×32 samples) — see Figure 3.3. The SF for the inverse quantisation process, denoted as s', is computed in (3.18):

DC $d = 0.0000$ $w = 1.0000$	AC $d = 0.1010$ $w = 0.9898$	AC $d = 0.2020$ $w = 0.9600$	AC $d = 0.3030$ $w = 0.9123$	AC $d = 0.4041$ $w = 0.8494$	AC $d = 0.5051$ $w = 0.7748$	AC $d = 0.6061$ $w = 0.6926$	AC $d = 0.7071$ $w = 0.6065$
AC $d = 0.1010$ $w = 0.9898$	AC $d = 0.1429$ $w = 0.9798$	AC $d = 0.2259$ $w = 0.9503$	AC $d = 0.3194$ $w = 0.9030$	AC $d = 0.4165$ $w = 0.8407$	AC $d = 0.5151$ $w = 0.7670$	AC $d = 0.6145$ $w = 0.6855$	AC $d = 0.7143$ $w = 0.6004$
AC $d = 0.2020$ $w = 0.9600$	AC $d = 0.2259$ $w = 0.9503$	AC $d = 0.2857$ $w = 0.9216$	AC $d = 0.3642$ $w = 0.8758$	AC $d = 0.4518$ $w = 0.8154$	AC $d = 0.5540$ $w = 0.7438$	AC $d = 0.6389$ $w = 0.6649$	AC $d = 0.7354$ $w = 0.5823$
AC $d = 0.3030$ $w = 0.9123$	AC $d = 0.3194$ $w = 0.9030$	AC $d = 0.3642$ $w = 0.8758$	AC $d = 0.4286$ $w = 0.8322$	AC $d = 0.5051$ $w = 0.7748$	AC $d = 0.5890$ $w = 07068$	AC $d = 0.6776$ $w = 0.6318$	AC $d = 0.7693$ $w = 0.5533$
AC $d = 0.4041$ $w = 0.8494$	AC $d = 0.4165$ $w = 0.8407$	AC $d = 0.4518$ $w = 0.8154$	AC $d = 0.5051$ $w = 0.7748$	AC $d = 0.5714$ $w = 0.7214$	AC $d = 0.6468$ $w = 0.6581$	AC $d = 0.7284$ $w = 0.5882$	AC $d = 0.8144$ $w = 0.5152$
AC $d = 0.5051$ $w = 0.7748$	AC $d = 0.5151$ $w = 0.7670$	AC $d = 0.5440$ $w = 0.7438$	AC $d = 0.5890$ $w = 0.7068$	AC $d = 0.6468$ $w = 0.6581$	AC $d = 0.7143$ $w = 0.6004$	AC $d = 0.7890$ $w = 0.5366$	AC $d = 0.8690$ $w = 0.4700$
AC $d = 0.6061$ $w = 0.6926$	AC $d = 0.6145$ $w = 0.6855$	AC $d = 0.6389$ $w = 0.6649$	AC $d = 0.6776$ $w = 0.6318$	AC $d = 0.7284$ $w = 0.5882$	AC $d = 0.7890$ $w = 0.5366$	AC $d = 0.8571$ $w = 0.4797$	AC $d = 0.9313$ $w = 0.4201$
AC $d = 0.7071$ $w = 0.6065$	AC $d = 0.7143$ $w = 0.6004$	AC $d = 0.7354$ $w = 0.5823$	AC $d = 0.7693$ $w = 0.5533$	AC $d = 0.8144$ $w = 0.5152$	AC $d = 0.8690$ $w = 0.4700$	AC $d = 0.9313$ $w = 0.4201$	AC $d = 1.0000$ $w = 0.3679$

Figure 3.3: A graphical representation of the normalised Euclidean distance d and weight w associated with transform coefficient locations in a 8×8 TB. The darker orange, lighter orange, yellow and grey squares correspond to the DC coefficient, low frequency AC coefficients, medium frequency AC coefficients and high frequency AC coefficients, respectively.

$$s' = \left(\frac{2^{20}}{m} \right) \cdot e^{-(d)^2} \tag{3.18}$$

Note that the values of m and w are available at the decoder side. MF m is known from the bitstream after entropy decoding with CABAC; moreover, distance parameter d is determined by the transform coefficient location in luma and chroma TBs. Therefore, the value for s' employed in the encoder loop for generating reference frames is the same value as s' at the decoder side. As such, this allows for an efficient and low complexity encoder side implementation of TCPQ; note that all of the values associated with the proposed method are signalled to the decoder in the PPS.

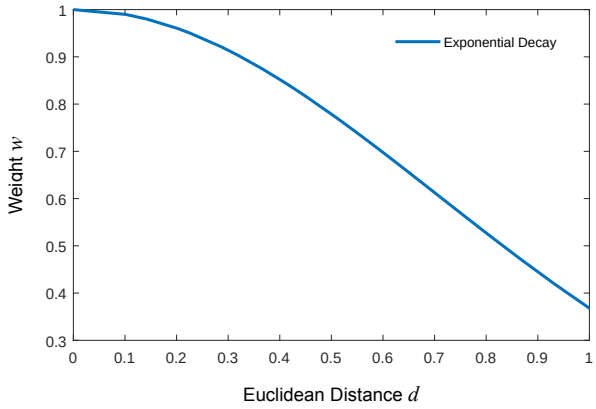


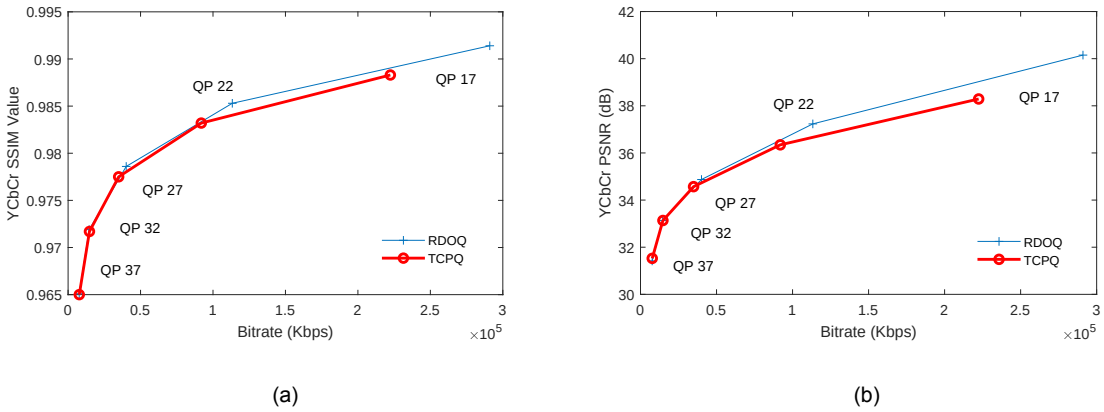
Figure 3.4: A plot showing the exponential decay in relation to weight *w* with respect to Euclidean distance *d*. Note *w* decreases as the distance increases. In addition, the curve corresponds closely to the MTF of the HVS.

In relation to the TCPQ, note that the quantisation and inverse quantisation procedures are identical for all TB sizes (i.e., from 4×4 samples to 32×32 samples). It is important to mention that weight *w*, by decreasing the MF, indirectly increases the QP value for each AC transform coefficient without the need to analyse multiple QPs (as is the case with RDOQ). Consequently, this can give rise to improvements in terms of computational complexity reductions and runtime decreases.

In addition to the primary objective of achieving perceptual quantisation (i.e., significant decreases in bitrates compared with RDOQ and URQ), another objective is to ensure that computational complexity and the associated encoding and decoding runtimes are not increased. As is the case with URQ, the computational complexity of TCPQ is computed in linear time *T*, as computed in (3.19):

$$T(n) = O(n) \tag{3.19}$$

Like URQ, the computational performance of TCPQ is directly proportional to the number of transform coefficients being processed in each luma and chroma TB. URQ in HEVC quantises transform coefficients individually in each TB by virtue of the computation in (3.7). Refining MF *m* in (3.7) with weight *w* in (3.16) does not increase computational complexity. Moreover, TCPQ reduces the number of non-zero quantised transform coefficients, which may give rise to faster entropy coding and decoding times compared with RDOQ. This is because fewer non-zero coefficients results in an encoded bitstream with fewer bits compared with RDOQ.



(a) (b)

Figure 3.5: Two plots which highlight the bitrate reductions attained by TCPQ compared with RDOQ. The subfigures show the bitrate reductions achieved by TCPQ on the following sequences. Subfigure (a): BirdsInCage 4:4:4 (AI - SSIM). Subfigure (b): BirdsInCage 4:4:4 (AI - PSNR).

3.3 Experimental Evaluations, Results and Discussion

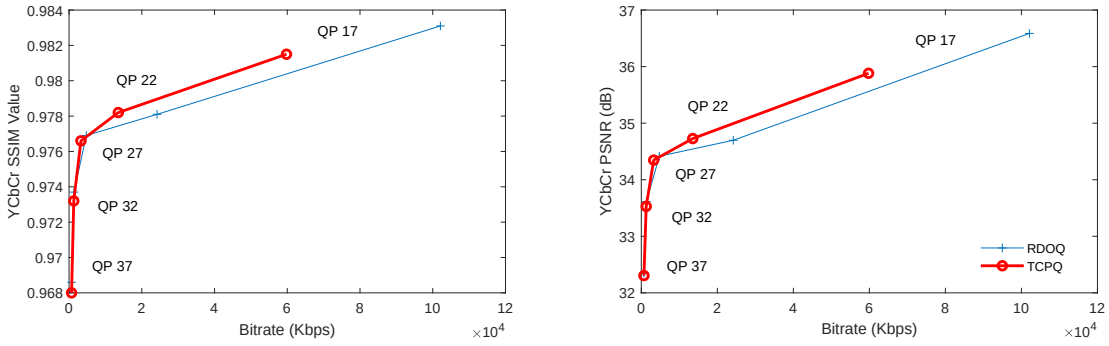
The same experimental evaluation procedure is employed in all three contribution chapters of this PhD thesis. Please refer to section 1.2 in Chapter 1 for a detailed overview of the conditions employed in the objective and subjective evaluations.

TCPQ is evaluated and compared with the RDOQ [35, 36], which is the default quantisation technique employed in HEVC. It is important to affirm that RDOQ, by definition, always significantly outperforms URQ in terms of coding efficiency [13]. Therefore, comparing TCPQ with RDOQ also implies comparing TCPQ with URQ.

In terms of the bitrate reductions achieved by TCPQ versus RDOQ, Table 3.2 shows that the highest bitrate reductions in the AI tests are attained on the BirdsInCage 4:4:4 and OldTownCross 4:4:4 sequences — see Figure 3.5. Similarly, the highest overall bitrate reductions accomplished (RA tests) by TCPQ are gained on the BirdsInCage 4:4:4 and OldTownCross 4:4:4 sequences with overall bitrate decrease of 41% and 34.7%, respectively. Note that, according to the subjective evaluations, no perceivable differences between TCPQ versus RDOQ were recorded in any of the tests undertaken. More specifically, over five QP data points (i.e., initial QPs 17, 22, 27, 32 and 37) using the AI and RA encoding configurations, high bitrate reductions are attained by TCPQ in addition to accomplishing perceptually identical reconstruction quality.

Table 3.2: Tabulated bitrate reduction percentages attained by the proposed TCPQ technique compared with RDOQ. The bitrate reductions are averaged over five QP data points (initial QPs 17, 22, 27, 37 and 37). The AI results are shown on the left; the RA results are shown on the right. The text in red indicates bitrate inflations.

TCPQ Versus RDOQ (YCbCr 4:2:0) – AI		**TCPQ Versus RDOQ (YCbCr 4:2:0) – RA**	
Sequence	Bitrate (%)	**Sequence**	Bitrate (%)
BirdsInCage	**−8.6**	BirdsInCage	**−30.1**
DuckAndLegs	**−10.8**	DuckAndLegs	**−25.3**
Kimono	**−4.7**	Kimono	**−8.2**
OldTownCross	**−14.0**	OldTownCross	**−30.0**
ParkScene	**−6.6**	ParkScene	**−9.9**
Traffic	**−3.9**	Traffic	**−9.3**
TCPQ Versus RDOQ (YCbCr 4:2:2) – AI		**TCPQ Versus RDOQ (YCbCr 4:2:2) – RA**	
Sequence	Bitrate (%)	**Sequence**	Bitrate (%)
BirdsInCage	**−11.2**	BirdsInCage	**−34.2**
DuckAndLegs	**−13.9**	DuckAndLegs	**−30.5**
Kimono	**−10.3**	Kimono	**−16.4**
OldTownCross	**−16.7**	OldTownCross	**−33.0**
ParkScene	**−17.2**	ParkScene	**−22.9**
Traffic	**−4.0**	Traffic	**−11.2**
TCPQ Versus RDOQ (YCbCr 4:4:4) – AI		**TCPQ Versus RDOQ (YCbCr 4:4:4) – RA**	
Sequence	Bitrate (%)	**Sequence**	Bitrate (%)
BirdsInCage	**−20.4**	BirdsInCage	**−41.0**
DuckAndLegs	**−17.0**	DuckAndLegs	**−32.4**
Kimono	**−15.4**	Kimono	**−23.7**
OldTownCross	**−19.7**	OldTownCross	**−34.7**
ParkScene	**−13.5**	ParkScene	**−26.2**
Traffic	**−4.7**	Traffic	**−14.3**



(a) (b)

Figure 3.6: Two plots which highlight the bitrate reductions attained by TCPQ compared with RDOQ. The subfigures show the bitrate reductions achieved by TCPQ on the following sequences. Subfigure (a): BirdsInCage 4:4:4 (RA - SSIM). Subfigure (b): BirdsInCage 4:4:4 (RA - PSNR).

(a) Luma Channel

(b) Chroma Cb Channel

(c) Chroma Cr Channel

Figure 3.7: The SSIM Index Map (structural reconstruction errors) of a TCPQ-coded inter-frame (RA QP = 22 test) versus the raw data (ParkScene 4:4:4 sequence). In subfigures (a), (b) and (c), respectively, the luma (Y), chroma Cb and chroma Cr structural reconstruction errors are shown separately. Note that these reconstruction errors in the TCPQ-coded compressed sequence are imperceptible to the HVS according to the subjective evaluations (compare the subfigures in Figure 3.8).

(a) TCPQ-Coded Inter-Frame (RA = QP 22): YCbCr PSNR = 32.8848 and YCbCr SSIM = 0.8629

(b) Raw Data

Figure 3.8: A frame from the ParkScene 4:4:4 sequence. Subfigure (a) is a TCPQ-coded inter-frame from this sequence (RA QP = 22 test). Subfigure (b) is the corresponding raw data. Note that the TCPQ-coded sequence in (a) is perceptually indistinguishable from the raw data in (b) according to the subjective evaluations.

Table 3.3: The SSIM results for the 'TCPQ versus the raw data' in comparison with 'RDOQ versus the raw data' (initial QPs 17, 22, 27, 32 and 37) using the AI encoding configuration. Green text indicates superior results.

Mean SSIM Values (Per Sequence, Per QP): TCPQ Versus RDOQ (YCbCr 4:2:0) – All Intra

Sequence	TCPQ					RDOQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	0.9925	0.9898	0.9863	0.9814	0.9760	**0.9932**	**0.9902**	**0.9865**	**0.9818**	**0.9761**
DuckAndLegs	0.9721	0.9480	0.9153	0.8826	0.8369	**0.9861**	**0.9611**	**0.9187**	**0.8856**	**0.8393**
Kimono	0.9569	0.9376	0.9187	0.8946	0.8644	**0.9619**	**0.9378**	**0.9189**	**0.8962**	**0.8664**
OldTownCross	0.9442	0.9025	0.8601	0.8290	**0.7863**	**0.9664**	**0.9152**	**0.8627**	**0.8297**	0.7856
ParkScene	0.9683	0.9509	0.9255	0.8873	0.8379	**0.9728**	**0.9557**	**0.9292**	**0.8901**	**0.8375**
Traffic	0.9769	0.9632	0.9395	0.9045	0.8551	**0.9799**	**0.9656**	**0.9418**	**0.9056**	**0.8542**

Mean SSIM Values (Per Sequence, Per QP): TCPQ Versus RDOQ (YCbCr 4:2:2) – All Intra

Sequence	TCPQ					RDOQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	0.9913	0.9881	0.9843	0.9787	0.9709	**0.9926**	**0.9888**	**0.9844**	**0.9788**	**0.9713**
DuckAndLegs	0.9698	0.9422	0.8985	0.8542	0.8014	**0.9877**	**0.9644**	**0.9066**	**0.8564**	**0.8028**
Kimono	0.9487	0.9213	0.8981	0.8711	0.8310	**0.9615**	**0.9236**	**0.8981**	**0.8718**	**0.8319**
OldTownCross	0.9386	0.8838	0.8299	0.7955	**0.7480**	**0.9682**	**0.9053**	**0.8300**	**0.7970**	0.7474
ParkScene	0.9407	0.9015	0.8635	0.8257	0.7774	**0.9626**	**0.9129**	**0.8654**	**0.8265**	**0.7775**
Traffic	0.9786	0.9650	0.9420	0.9032	**0.8407**	**0.9816**	**0.9673**	**0.9437**	**0.9044**	0.8391

Mean SSIM Values (Per Sequence, Per QP): TCPQ Versus RDOQ (YCbCr 4:4:4) – All Intra

Sequence	TCPQ					RDOQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	0.9883	0.9832	0.9775	0.9717	0.9650	**0.9914**	**0.9853**	**0.9786**	**0.9719**	**0.9651**
DuckAndLegs	0.9718	0.9473	0.8949	0.8301	0.7751	**0.9900**	**0.9718**	**0.9201**	**0.8338**	**0.7771**
Kimono	0.9494	0.9145	**0.8793**	0.8512	0.8146	**0.9653**	**0.9245**	0.8786	**0.8533**	**0.8161**
OldTownCross	0.9425	0.8810	0.7882	0.7318	**0.6861**	**0.9745**	**0.9198**	**0.7978**	**0.7330**	0.6860
ParkScene	0.9469	0.9047	0.8547	0.8134	0.7663	**0.9669**	**0.9205**	**0.8571**	**0.8146**	**0.7669**
Traffic	0.9798	0.9657	0.9425	0.9063	0.8483	**0.9828**	**0.9684**	**0.9452**	**0.9082**	**0.8485**

The subfigures in Figure 3.7 highlight the structural luma and chroma reconstruction errors in the 'TCPQ versus the raw data' test conducted on the ParkScene 4:4:4 sequence. In comparison with the raw data, the structural reconstruction errors incurred by TCPQ in the Y, Cb and Cr channels are concentrated mostly in the high variance regions in each channel. In spite of these quantisation-induced errors in the TCPQ-coded sequence, visually lossless coding is achieved in the AI and RA QP = 22 tests, as confirmed in the subjective evaluations. Compare the subfigures in Figure 3.8 for a visual example.

Table 3.4: The SSIM results for the 'TCPQ versus the raw data' in comparison with 'RDOQ versus the raw data' (initial QPs 17, 22, 27, 32 and 37) using the RA encoding configuration. Green text indicates superior results.

Mean SSIM Values (Per Sequence, Per QP): TCPQ Versus RDOQ (YCbCr 4:2:0) – Random Access

Sequence	TCPQ					RDOQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	0.9904	0.9894	0.9872	0.9836	0.9797	**0.9905**	**0.9897**	**0.9877**	**0.9843**	**0.9801**
DuckAndLegs	0.9371	0.9173	0.9001	0.8734	0.8320	**0.9415**	**0.9208**	**0.9055**	**0.8791**	**0.8384**
Kimono	0.9320	0.9236	0.9087	0.8874	0.8640	**0.9324**	**0.9253**	**0.9111**	**0.8912**	**0.8668**
OldTownCross	0.8856	0.8641	0.8531	0.8347	0.8001	**0.8890**	**0.8648**	**0.8559**	**0.8386**	**0.8047**
ParkScene	0.9587	0.9450	0.9213	0.8859	0.8396	**0.9629**	**0.9505**	**0.9278**	**0.8936**	**0.8460**
Traffic	0.9719	0.9619	0.9429	0.9132	0.8695	**0.9746**	**0.9652**	**0.9468**	**0.9174**	**0.8745**

Mean SSIM Values (Per Sequence, Per QP): TCPQ Versus RDOQ (YCbCr 4:2:2) – Random Access

Sequence	TCPQ					RDOQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	0.9880	**0.9867**	0.9848	0.9811	0.9755	**0.9882**	**0.9867**	**0.9850**	**0.9813**	**0.9759**
DuckAndLegs	0.9324	0.8939	0.8763	0.8465	0.8002	**0.9463**	**0.8965**	**0.8819**	**0.8532**	**0.8074**
Kimono	0.9085	**0.9001**	0.8873	0.8667	0.8372	**0.9072**	**0.9001**	**0.8891**	**0.8693**	**0.8408**
OldTownCross	0.8532	**0.8222**	0.8136	0.7955	0.7567	**0.8541**	0.8221	**0.8161**	**0.7987**	**0.7606**
ParkScene	0.8934	0.8797	0.8603	0.8299	0.7863	**0.8972**	**0.8835**	**0.8653**	**0.8339**	**0.7908**
Traffic	0.9713	0.9614	0.9436	0.9122	0.8585	**0.9740**	**0.9648**	**0.9476**	**0.9165**	**0.8660**

Mean SSIM Values (Per Sequence, Per QP): TCPQ Versus RDOQ (YCbCr 4:4:4) – Random Access

Sequence	TCPQ					RDOQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	0.9815	**0.9782**	0.9766	0.9732	0.9680	**0.9831**	0.9781	**0.9769**	**0.9737**	**0.9686**
DuckAndLegs	0.9377	0.8853	0.8477	0.8181	0.7697	**0.9610**	**0.8944**	**0.8516**	**0.8234**	**0.7769**
Kimono	0.8994	**0.8740**	0.8629	0.8446	0.8172	**0.9008**	0.8736	**0.8646**	**0.8474**	**0.8218**
OldTownCross	0.8465	**0.7512**	0.7457	0.7292	0.6937	**0.8715**	0.7485	**0.7477**	**0.7328**	**0.6986**
ParkScene	0.8959	0.8629	0.8451	0.8167	0.7739	**0.9016**	**0.8644**	**0.8492**	**0.8214**	**0.7789**
Traffic	0.9697	0.9599	0.9426	0.9134	0.8641	**0.9725**	**0.9633**	**0.9468**	**0.9179**	**0.8705**

In terms of the mathematical reconstruction quality of the compressed video data, as shown in Table 3.3 and Table 3.4 the SSIM values for the TCPQ-coded sequences are not significantly different from those of the RDOQ-coded sequences. In most cases, the SSIM values for the RDOQ-coded sequences are higher. However, these results do not reflect how the subjective evaluation participants perceived the visual quality of the compressed video sequences. That is, the differences in mathematical reconstruction quality, as quantified by SSIM, are perceptually indiscernible according to the subjective evaluations conducted.

Table 3.5: The PSNR results for the 'TCPQ versus the raw data' in comparison with 'RDOQ versus the raw data' (initial QPs 17, 22, 27, 32 and 37) using the AI encoding configuration. Green text indicates superior results.

Mean PSNR (dB) Per Sequence, Per QP: TCPQ Versus RDOQ (YCbCr 4:2:0) – All Intra

Sequence	TCPQ					RDOQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	40.2296	38.4002	36.6354	34.7860	**32.8728**	**41.0129**	**38.7012**	**36.6945**	**34.7871**	32.7339
DuckAndLegs	36.1747	33.5289	31.0645	29.0790	27.1250	**39.3833**	**34.8995**	**31.4116**	**29.2380**	**27.1996**
Kimono	39.1947	37.3183	35.5245	33.6365	31.6790	**39.6659**	**37.3840**	**35.5691**	**33.7496**	**31.7134**
OldTownCross	36.8165	34.3910	32.4754	30.9350	**29.2072**	**39.1555**	**35.1321**	**32.6708**	**31.0068**	29.1971
ParkScene	38.3919	35.9489	33.5807	31.2039	28.9970	**40.2817**	**37.0864**	**34.0211**	**31.3811**	**29.0183**
Traffic	39.0849	36.7575	34.1328	31.6338	29.2000	**40.5629**	**37.5261**	**34.5096**	**31.7751**	**29.2048**

Mean PSNR (dB) Per Sequence, Per QP: TCPQ Versus RDOQ (YCbCr 4:2:2) – All Intra

Sequence	TCPQ					RDOQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	39.5318	37.7115	36.0440	**34.2490**	**32.1414**	**40.6194**	**38.1548**	**36.0751**	34.2352	32.0200
DuckAndLegs	35.5482	32.8397	30.3098	28.2637	26.2842	**39.4249**	**34.8383**	**30.7462**	**28.3535**	**26.3148**
Kimono	38.3566	36.4261	34.8801	33.0928	30.9228	**39.4174**	**36.5508**	**34.8996**	**33.1495**	**30.9463**
OldTownCross	36.2502	33.6149	31.7518	30.3369	**28.6096**	**39.1385**	**34.5342**	**31.8470**	**30.3877**	28.5875
ParkScene	36.8270	34.5044	32.4421	30.4694	**28.5484**	**39.1704**	**35.3941**	**32.7094**	**30.5337**	28.5243
Traffic	39.3220	36.9278	34.2833	31.5394	**28.8084**	**40.9430**	**37.7508**	**34.6483**	**31.6633**	28.7790

Mean PSNR (dB) Per Sequence, Per QP: TCPQ Versus RDOQ (YCbCr 4:4:4) – All Intra

Sequence	TCPQ					RDOQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	38.2903	36.3439	34.5694	33.1340	**31.5270**	**40.1517**	**37.2325**	**34.8728**	**33.1364**	31.4187
DuckAndLegs	35.2366	32.6017	29.7395	27.6077	25.8644	**39.6440**	**35.1159**	**30.6737**	**27.7083**	**25.9131**
Kimono	38.0461	35.8240	34.1624	32.6565	30.7349	**39.4838**	**36.2514**	**34.1785**	**32.7523**	**30.7730**
OldTownCross	35.8234	32.8982	30.5346	29.1914	27.8360	**39.3534**	**34.4861**	**30.7330**	**29.2415**	**27.8500**
ParkScene	37.1236	34.6297	32.4205	30.5104	28.5994	**39.3890**	**35.5348**	**32.6720**	**30.6046**	**28.6032**
Traffic	39.5682	37.0596	34.3502	31.6235	28.8940	**41.2753**	**37.9611**	**34.7950**	**31.8105**	**28.9053**

The proposed TCPQ method and RDOQ operate in a comparable manner. Therefore, the structural reconstruction quality is relatively similar in both TCPQ-coded sequences and RDOQ-coded sequences. Recall that RDOQ establishes a suitable trade-off between the bitrate and compression-induced distortion; this is achieved by minimising the rate-distortion Lagrangian cost function. Consequently, higher levels of quantisation are applied to high frequency AC coefficients in luma and chroma TBs. Similarly, TCPQ, with the proposed Euclidean distance parameter approach, ensures that high frequency AC coefficients are quantised to a much higher degree than the DC coefficient and the low frequency AC coefficients.

Table 3.6: The PSNR results for the 'TCPQ versus the raw data' in comparison with 'RDOQ versus the raw data' (initial QPs 17, 22, 27, 32 and 37) using the RA encoding configuration. Green text indicates superior results.

Mean PSNR (dB) Per Sequence, Per QP: TCPQ Versus RDOQ (YCbCr 4:2:0) – Random Access

Sequence	TCPQ					RDOQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	38.6175	38.0313	36.9060	35.4596	33.8542	**38.7047**	**38.1957**	**37.1156**	**35.6208**	**33.9313**
DuckAndLegs	32.6849	31.1758	29.8329	28.3225	26.6617	**33.0892**	**31.5517**	**30.2455**	**28.6183**	**26.8910**
Kimono	37.1279	36.0300	34.4623	32.6953	31.0570	**37.1744**	**36.1917**	**34.6534**	**32.9357**	**31.2515**
OldTownCross	33.8608	32.8185	32.1905	31.2190	29.8049	**34.1082**	**32.8888**	**32.3525**	**31.4017**	**29.9488**
ParkScene	37.1455	35.2638	33.1639	31.0835	29.1460	**38.1238**	**36.1143**	**33.7870**	**31.5525**	**29.4531**
Traffic	37.9109	36.1445	34.0528	31.8976	29.7047	**38.6447**	**36.8452**	**34.5873**	**32.2633**	**29.9602**

Mean PSNR (dB) Per Sequence, Per QP: TCPQ Versus RDOQ (YCbCr 4:2:2) – Random Access

Sequence	TCPQ					RDOQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	37.6014	36.9339	36.1972	34.8992	33.1689	**37.7717**	**36.9658**	**36.3165**	**34.9791**	**33.2516**
DuckAndLegs	32.0353	30.1167	29.0295	27.5933	25.9268	**32.8429**	**30.3252**	**29.3686**	**27.8630**	**26.1294**
Kimono	**35.8131**	35.0102	33.7945	**32.2135**	30.5026	35.7653	**35.0352**	**33.9246**	**32.2135**	**30.7039**
OldTownCross	32.8684	31.8850	31.4376	30.6083	29.2096	**32.9684**	**31.8983**	**31.5668**	**30.7429**	**29.3323**
ParkScene	34.3273	33.2331	31.9147	30.3856	28.7054	**34.6558**	**33.6197**	**32.3031**	**30.6253**	**28.8734**
Traffic	37.7826	36.0363	34.0288	31.7721	29.3266	**38.5194**	**36.7086**	**34.5679**	**32.1334**	**29.6176**

Mean PSNR (dB) Per Sequence, Per QP: TCPQ Versus RDOQ (YCbCr 4:4:4) – Random Access

Sequence	TCPQ					RDOQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	35.8818	**34.7279**	34.3471	33.5290	32.3055	**36.5869**	34.6978	**34.4153**	**33.6150**	**32.3482**
DuckAndLegs	31.5745	29.0506	27.8146	26.6442	25.1170	**33.3669**	**29.3025**	**27.9705**	**26.8144**	**25.2824**
Kimono	35.1706	33.9955	33.0732	31.7813	30.2418	**35.2126**	**33.9887**	**33.1850**	**31.9445**	**30.4645**
OldTownCross	32.0615	**30.1570**	29.9503	29.3844	28.3097	**32.7350**	30.0759	**30.0171**	**29.4926**	**28.4508**
ParkScene	34.3445	32.8848	31.7675	30.3512	28.6899	**34.6814**	**33.0917**	**32.0614**	**30.6221**	**28.8909**
Traffic	37.6342	35.8630	33.9318	31.7879	29.4056	**38.3907**	**36.5171**	**34.4577**	**32.1694**	**29.7062**

Overall, the objective reconstruction quality of the RDOQ-coded sequences, as quantified by SSIM and PSNR (see Table 3.3 to Table 3.6), proved to be superior in the vast majority of cases. However, according to the MOS results obtained via the subjective evaluations, the participants did not notice any perceivable visual quality differences between any of the TCPQ-coded sequences and the RDOQ-coded sequences (see Table 3.7). This provides evidence that comparatively high SSIM and PSNR values do not necessarily equate to superior perceptual quality. Furthermore, this observation pertains to the fact that subjective evaluations are critically important in terms of robustly assessing HVS-orientated perceptual coding techniques.

Table 3.7: The MOS results, rounded to the nearest integer, of four participants in the subjective evaluations for TCPQ versus RDOQ.

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – TCPQ versus RDOQ

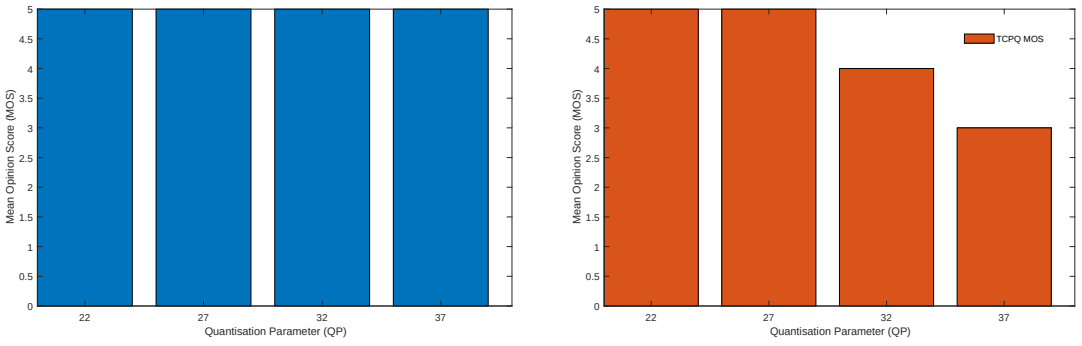
Sequence	YCbCr 4:2:0 All Intra				YCbCr 4:2:0 Random Access			
	QP 22	**QP 27**	**QP 32**	**QP 37**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	5	5	5	5	5	5	5	5
DuckAndLegs	5	5	5	5	5	5	5	5
Kimono	5	5	5	5	5	5	5	5
OldTownCross	5	5	5	5	5	5	5	5
ParkScene	5	5	5	5	5	5	5	5
Traffic	5	5	5	5	5	5	5	5

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – TCPQ versus RDOQ

Sequence	YCbCr 4:2:2 All Intra				YCbCr 4:2:2 Random Access			
	QP 22	**QP 27**	**QP 32**	**QP 37**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	5	5	5	5	5	5	5	5
DuckAndLegs	5	5	5	5	5	5	5	5
Kimono	5	5	5	5	5	5	5	5
OldTownCross	5	5	5	5	5	5	5	5
ParkScene	5	5	5	5	5	5	5	5
Traffic	5	5	5	5	5	5	5	5

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – TCPQ versus RDOQ

Sequence	YCbCr 4:4:4 All Intra				YCbCr 4:4:4 Random Access			
	QP 22	**QP 27**	**QP 32**	**QP 37**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	5	5	5	5	5	5	5	5
DuckAndLegs	5	5	5	5	5	5	5	5
Kimono	5	5	5	5	5	5	5	5
OldTownCross	5	5	5	5	5	5	5	5
ParkScene	5	5	5	5	5	5	5	5
Traffic	5	5	5	5	5	5	5	5



(a) (b)

Figure 3.9: Two Mean Opinion Score (MOS) bar graphs. Subfigure (a) shows the MOS for TCPQ versus RDOQ on all sequences using the AI and RA configurations. Subfigure (b) shows the MOS for TCPQ versus the raw video data on the BirdsInCage 4:4:4 10-bit sequence.

Table 3.8: The MOS results, rounded to the nearest integer, of four participants in the subjective evaluations for TCPQ versus the raw video data.

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – TCPQ versus Raw Data

Sequence	YCbCr 4:2:0 All Intra				YCbCr 4:2:0 Random Access			
	QP 22	QP 27	QP 32	QP 37	QP 22	QP 27	QP 32	QP 37
BirdsInCage	5	4	2	1	5	4	3	2
DuckAndLegs	5	4	3	1	5	5	4	3
Kimono	5	3	2	1	5	4	3	2
OldTownCross	5	3	1	1	5	4	3	2
ParkScene	5	4	2	1	5	4	3	2
Traffic	5	4	2	1	5	5	4	3

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – TCPQ versus Raw Data

Sequence	YCbCr 4:2:2 All Intra				YCbCr 4:2:2 Random Access			
	QP 22	QP 27	QP 32	QP 37	QP 22	QP 27	QP 32	QP 37
BirdsInCage	5	4	2	1	5	4	3	2
DuckAndLegs	5	4	3	1	5	5	3	2
Kimono	5	4	2	1	5	4	3	2
OldTownCross	5	4	2	1	5	4	3	2
ParkScene	5	4	2	1	5	5	4	3
Traffic	5	5	2	1	5	5	4	3

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – TCPQ versus Raw Data

Sequence	YCbCr 4:4:4 All Intra				YCbCr 4:4:4 Random Access			
	QP 22	QP 27	QP 32	QP 37	QP 22	QP 27	QP 32	QP 37
BirdsInCage	5	4	3	1	5	5	4	3
DuckAndLegs	5	5	4	2	5	5	4	3
Kimono	5	4	3	1	5	4	3	2
OldTownCross	5	4	2	1	5	4	3	2
ParkScene	5	4	3	1	5	5	4	3
Traffic	5	4	2	1	5	5	4	3

Table 3.7 and Table 3.8 tabulate the rounded MOS for the four subjective evaluation participants. Table 3.7 includes the MOS results for ‘TCPQ versus RDOQ’ and Table 3.8 shows the MOS results for ‘TCPQ versus the raw video data’. As shown in Table 3.7, the subjective evaluation participants were unable to detect any perceptually discernible differences between the TCPQ-coded sequences and the RDOQ-coded sequences. An MOS value of 5 is recorded for all tests on all sequences (i.e., the AI and RA tests using initial QPs 17, 22, 27, 32 and 37 on the 4:2:0, 4:2:2 and 4:4:4 versions of each sequence).

The MOS results tabulated in Table 3.8 are significantly different from those shown in Table 3.7. Visually lossless coding is achieved by TCPQ in all of the RA QP = 22 tests. Similarly, in almost all of the AI QP = 22 tests, visually lossless coding is accomplished by TCPQ. It is important to note that, in certain QP = 27 tests including the RA QP = 27 tests conducted on the BirdsInCage 4:4:4 and DuckAndLegs 4:4:4 sequences, visually lossless coding is attained by TCPQ; this is significant from a bitrate reduction perspective. Recall from Table 3.2 and compared with RDOQ, TCPQ attains 39.4% and 32.8% bitrate reductions when applied to the BirdsInCage 4:4:4 sequence and the DuckAndLegs 4:4:4 sequence, respectively.

In the vast majority of cases, the TCPQ-coded sequences using the RA encoding configuration (i.e., the RA GOP-based inter coding tests) were perceived to be vastly superior compared with the sequences coded using the All Intra configuration. This is because motion data with GOP-based inter coding in HEVC can be signalled with the utilisation of merge mode or by motion vector differences, picture reference indices and the direction of the inter prediction [10, 11].

Recall that the 4:4:4 (and 4:2:2) versions of the BirdsInCage sequence and the DuckAndLegs sequence are high bit depth 30-bit sequences (i.e., 10-bits per channel). Note that 30-bit video data contains a much larger number of colours per pixel compared with 24-bit video data (i.e., potentially up to 1024^3 colours per pixel). Therefore, in combination with the absence of chroma subsampling in YCbCr 4:4:4 data, the high bit depth characteristics of these 30-bit sequences equates to the fact that the 10-bit Y, Cb and Cr channels comprise higher variances compared with the 8-bit Y, Cb and Cr channels in 24-bit YCbCr 4:2:0 chroma subsampled video data. To reiterate, it has been established in the literature that it is more difficult for the HVS to detect compression-induced artifacts in high variance regions of image and video data, which constitutes high variance-based spatial masking [37, 39, 64]. This visual masking phenomenon is more prominent in high bit depth 4:4:4 data [23]. With this in mind, discarding high frequency detail in high variance luma and chroma data (i.e., 30-bit YCbCr 4:4:4 sequences) is typically not noticeable to the HVS.

In accordance with the exhaustive experimental evaluations conducted in this chapter, we can assert, with confidence, that visually lossless coding can be achieved without taking into account HVS-based models including spatial CSF-based spatiotemporal masking. Although TCPQ is a HVS-orientated technique which consists of perceptual considerations (i.e., quantising high frequency AC coefficient more coarsely than the DC coefficient and the low frequency AC coefficient), TCPQ is based on a conceptually simple adaptive Euclidean distance parameter.

It is important to note that both TCPQ and RDOQ are not suitable for low bitrate All Intra coding applications, as confirmed in the subjective evaluation results for the AI QP = 37 tests (see Table 3.8). This is due to the fact that both TCPQ and RDOQ are designed, for the most part, to preserve the integrity of the DC transform coefficient and low frequency AC coefficients in luma and chroma TBs. Moreover, in accordance with the subjective evaluation results, the quantisation-induced compression artifacts incurred by TCPQ in the RA QP = 37 tests are considerably less conspicuous than those that were induced in AI QP = 37 tests. To reiterate, this is because GOP-based inter coding includes the signalling of important motion data in the bitstream; All Intra coding does not account for motion data or the spatiotemporal redundancies that exist between frames. Therefore, the visual quality of the reconstructed inter-coded sequences — for both TCPQ and RDOQ — is significantly superior compared with the corresponding intra-coded sequences.

3.4 Summary

In this chapter, a coefficient-level perceptual quantisation technique, named TCPQ, is proposed. Compared with the most prominent and widely employed coefficient-level perceptual quantisation technique in video coding (named RDOQ), TCPQ achieves considerable bitrate reductions of up to 41% without incurring a discernible decrease of visual quality in the compressed video data. In the evaluations conducted, TCPQ proved to be very effective on high variance and high bit depth 4:4:4 sequences. In terms of runtimes, no significant differences are observed; slight reductions in encoding times and decoding times are achieved by TCPQ.

Chapter 4. Cross-Colour Channel and CB-Level Perceptual Quantisation

In this chapter, two novel HVS-based luma and chroma perceptual quantisation techniques are proposed. The motivation behind the development of the techniques in this chapter is the research gap in terms of HVS-based luminance and chrominance spatial masking and perceptual quantisation at the picture level in HEVC. The proposed TCPQ technique in Chapter 3 is a purely mathematical approach to transform coefficient-level perceptual quantisation based on a Euclidean distance parameter. In this chapter, the proposed techniques exploit HVS-based psychovisual masking in the spatial domain. The proposed methods are designed to improve upon the default perceptual quantisation technique in HEVC (AdaptiveQP [39]-[41]).

AdaptiveQP is a HVS-based perceptual quantisation method. It perceptually adjusts the QP at the $2N \times 2N$ CU level based on the spatial activity of the luminance data only. Because AdaptiveQP computes the spatial activity in luma CBs only, this means that the data in the chroma Cb and Cr CBs is quantised according to the spatial activity of luma samples only. By definition, the spatial activity of chroma Cb and Cr data is neglected in the AdaptiveQP technique, which constitutes a significant shortcoming. The proposed methods are designed to address this problem.

The first novel technique proposed in this chapter is a CU-level cross-colour channel perceptual quantisation method (CCCPQ) [22]. CCCPQ is designed to compute the combined spatial activity of the data in all CBs in a $2N \times 2N$ CU. The second novel technique proposed in this chapter is a CB-level perceptual quantisation method (CBPQ) [23]. CBPQ is suitable for high bit depth 4:4:4 video data because of the lack of chroma subsampling and also the potential high variance in the chroma channels. Compared with AdaptiveQP, CCCPQ and CBPQ are designed to significantly reduce bitrates in addition to the objective of achieving visually lossless coding.

The rest of the chapter is organised as follows. Section 4.1 includes background information relevant to the proposed methods. CCCPQ is presented in section 4.2, which includes the evaluation, results and discussion. CBPQ is presented in section 4.3, which includes the evaluation, results and discussion. Finally, section 4.4 concludes this chapter.

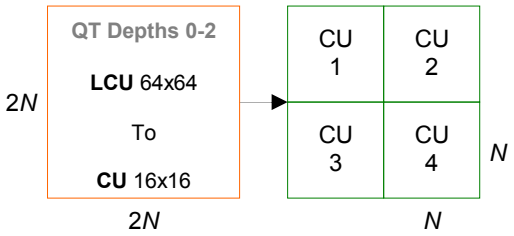


Figure 4.1: The CU size for which the QP is modified is 2*N*×2*N*. Both AdaptiveQP and CCCPQ operate at QT depth levels 0-2. When the split flag is enabled in HM, the 2*N*×2*N* CUs at QT depth levels 0-2 are partitioned into four constituent *N*×*N* CUs, where *N*=32 (level 0), *N*=16 (level 1) or *N*=8 (level 2). Note that CUs are always size 2*N*×2*N* or *N*×*N*. In other words and in contrast to CBs, CUs do not change in size due to chroma Cb and chroma Cr subsampling.

4.1 Related Background

The literature indicates that the HVS is significantly less sensitive to compression related distortions within regions of image data that comprise high pixel variations [37, 39, 64]. This is known as variance-based visual masking in the spatial domain. In the context of video coding and also the techniques proposed in this chapter, this equates to the fact that higher levels of quantisation can be applied to high spatial activity regions in all three colour channels of YCbCr video data. HVS-based perceptual quantisers can be designed for HEVC to exploit the variance-based spatial masking phenomenon of the HVS. That is, coarser levels of quantisation can be applied, at the CU and CB levels in HEVC, to regions in YCbCr video data that comprise high levels of spatial activity. As such, visual masking-based perceptual quantisation can potentially give rise to important bitrate reductions without incurring a perceptually discernible decrease of visual quality in the compressed video data.

Firstly, it is appropriate to distinguish the 2*N*×2*N* CU, the *N*×*N* CU and the CB. Assuming that the split flag is enabled in the HEVC HM reference software, the 2*N*×2*N* CU comprises four constituent *N*×*N* CUs (see Figure 4.1). The LCU supports 64×64 samples and the SCU supports 8×8 samples. LCUs operate at QuadTree (QT) Depth Level = 0 and SCUs operate at QT Depth Level = 3 [85, 14]. AdaptiveQP does not operate below QT Depth Level = 2. The CU, at all QT depth levels, comprises three CBs (assuming that the raw video data is not monochrome): one Y CB, one Cb CB and one Cr CB.

4.1.1 AdaptiveQP in HEVC

Recall that AdaptiveQP is a luma-only CU-level perceptual quantisation technique, which was officially adopted by JCT-VC for HEVC. AdaptiveQP perceptually adjusts the QP of an entire $2N\times2N$ CU based on the spatial activity of the sample data in the four constituent $N\times N$ SBs of the $2N\times2N$ luma CB only. Moreover, the spatial activity is quantified based on the variance of the sample values in the Sub-Blocks (SBs) of the luma CB. Therefore, a higher QP value is applied to a $2N\times2N$ CU in which the luma variance is high (due to the aforementioned HVS masking effect). Conversely, a lower QP value is applied to a $2N\times2N$ CU in which the luma variance is low. The principle drawback of AdaptiveQP is that it modifies the QP for chroma Cb and Cr CBs based on the variance of luma data only. This is not desirable because luma data is considerably different from chroma Cb and Cr data.

As an alternative quantisation technique to URQ in HEVC, AdaptiveQP is designed to exploit the phenomenon of HVS-based spatial masking by applying a higher QP — relative to the frame-level QP — to regions in a $2N\times2N$ CU in which the constituent luma CB consists of high variance of luma samples; this typically results in bitrate reductions compared with URQ. The fact that AdaptiveQP disregards the data in chroma Cb and Cr CBs during the CU QP selection process constitutes a significant shortcoming of this perceptual quantisation method. The contributions proposed in this chapter are designed to address this important drawback.

In terms of the mechanism of AdaptiveQP in HEVC HM, the CU-level QP, denoted as Q, is computed in (4.1):

$$Q = QP + \left[6 \times \log_2 \left(n \right) \right] \tag{4.1}$$

where QP corresponds to the frame-level QP and where n refers to the normalised spatial activity of samples in a luma CB. Variable n is computed in (4.2):

$$n = \frac{f \cdot l + t}{l + f \cdot t} \tag{4.2}$$

where f is a scaling factor associated with the CU-level QP adaptation range (denoted as a) regardless of the YCbCr colour channel; note that $a = 6$ is the default value in HEVC HM. Variable l corresponds to the spatial activity of samples in a luma CB and variable t refers to the mean spatial activity for all $2N \times 2N$ CUs. Variables f and l are computed in (4.3) and (4.4), respectively:

$$f = 2^{\frac{a}{6}} \tag{4.3}$$

$$l = 1 + \min\left(\sigma^2{}_{Y,k}\right), \quad \text{where } k = 1,....,4 \tag{4.4}$$

where $\sigma^2{}_{Y,k}$ denotes the spatial activity of samples in SB k (of size $N \times N$) of the luma CB. Variable $\sigma^2{}_{Y,k}$ is quantified as the variance of luma sample values, which is computed in (4.5):

$$\sigma^2{}_{Y,k} = \frac{1}{z}\sum_{i=1}^{z}\left(w_i - \mu_Y\right)^2 \tag{4.5}$$

where z denotes the number of luma samples in SB k of the luma CB. Variable w_i corresponds to the i^{th} luma sample in SB k and where μ_Y refers to the mean luma sample intensity of SB k, which is computed in (4.6).

$$\mu_Y = \frac{1}{z}\sum_{i=1}^{z} w_i \tag{4.6}$$

4.2 Proposed Cross-Colour Channel Perceptual Quantisation Technique

In CCCPQ [22], the proposed method achieves its objective by perceptually adjusting the QP, at the CU level, according to the combined variances of raw luma and chroma sample data in the constituent luma and chroma Cb and Cr CBs within a $2N \times 2N$ CU. Due to the fact that CCCPQ computes the combined spatial activity across all three colour channels (i.e., the Y, Cb and Cr channels), this equates to cross-colour channel dependency for QP selection; CCCPQ is particularly useful for perceptually quantising YCbCr 4:4:4 and YCbCr chroma subsampled data (i.e., 4:2:2 and 4:2:0).

In CCCPQ, the selection of the 2*N*×2*N* CU-level QP is contingent upon the spatial activity of the data in all three CBs, which constitutes cross-colour channel dependency for QP selection. CCCPQ adjusts the 2*N*×2*N* CU-level QP according to the sum of variances of the data in each of the four constituent *N*×*N* SBs of the luma CB and also the SBs of the chroma Cb and Cr CBs.

Like AdaptiveQP, CCCPQ does not operate below QT Depth Level = 2 (see Figure 4.1). By accounting for the raw luma and chroma sample data in luma CBs, chroma Cb CBs and chroma Cr CBs, respectively, CCCPQ is designed to derive a more appropriate QP selection that AdaptiveQP for the 2*N*×2*N* CU as a whole.

In CCCPQ, the CU-level QP, denoted as $\tilde{Q}$, is computed in (4.7):

$$\tilde{Q} = QP + \left[6 \times \log_2 \left(\tilde{n} \right) \right] \tag{4.7}$$

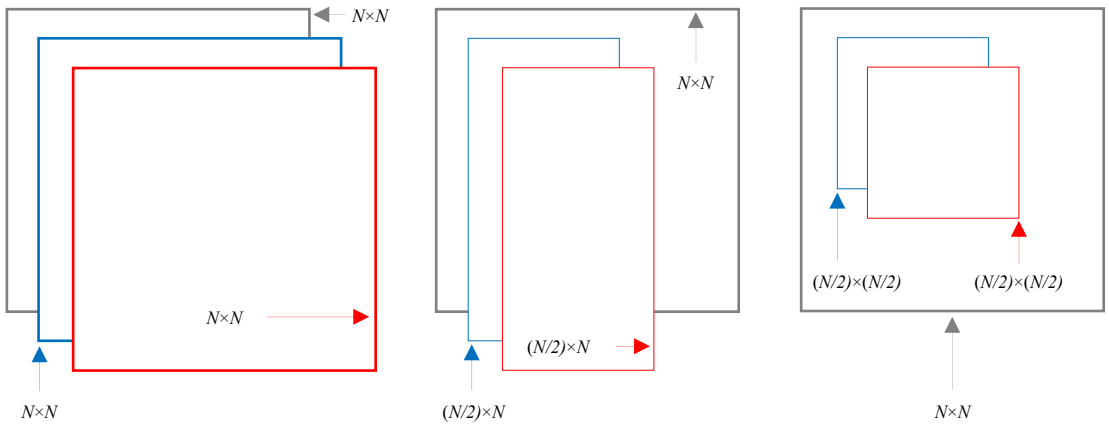
where $\tilde{n}$ denotes the normalised spatial activity of samples in both luma and chroma CBs. Variable $\tilde{n}$ is computed in (4.8):

$$\tilde{n} = \frac{f \cdot \left(l + b + d \right) + t}{\left(l + b + d \right) + f \cdot t} \tag{4.8}$$

where variables *b* and *d* correspond to the spatial activity of chroma Cb and Cr sample values in the chroma Cb and Cr CBs, respectively. Variables *b* and *d* are computed in (4.9) and (4.10), respectively.

$$b = 1 + \min \left(\sigma^2{}_{Cb,k} \right), \quad \text{where } k = 1,....,4 \tag{4.9}$$

$$d = 1 + \min \left(\sigma^2{}_{Cr,k} \right), \quad \text{where } k = 1,....,4 \tag{4.10}$$



(a) 4:4:4 YCbCr CB Sizes **(b)** 4:2:2 YCbCr CB Sizes **(c)** 4:2:0 YCbCr CB Sizes

Figure 4.2: The sizes of Y, Cb and Cr CB SBs in a $2N \times 2N$ CU in CCCPQ: Y (grey), Cb (blue), Cr (red). Each subfigure specifies the sizes of Cb and Cr SBs for different types of raw video data in terms of chroma sampling: (a) for 4:4:4 YCbCr video data, the CB SB sizes for Y, Cb and Cr are all $N \times N$, (b) for YCbCr 4:2:2 video data, the CB sizes are: $Y_{SB} = N \times N$, $Cb_{SB} = (N/2) \times N$ and $Cr_{SB} = (N/2) \times 2N$, and (c) for YCbCr 4:2:0 video data, the CB sizes are: $Y_{CB} = N \times N$, $Cb_{CB} = (N/2) \times (N/2)$ and $Cr_{CB} = (N/2) \times (N/2)$.

where $\sigma^2_{Cb,k}$ and $\sigma^2_{Cr,k}$ correspond to the spatial activity of chroma Cb and Cr data in SBs k of the chroma Cb and Cr CBs, respectively; note that Cb and Cr CBs — including the corresponding SBs — are not a fixed size, as shown in Figure 4.2. Variables $\sigma^2_{Cb,k}$ and $\sigma^2_{Cr,k}$ are computed as the variance of Cb and Cr sample values, respectively, as given by (4.11) and (4.12), respectively:

$$\sigma^2_{Cb,k} = \frac{1}{m} \sum_{i=1}^{m} \left(v_i - \mu_{Cb} \right)^2 \tag{4.11}$$

$$\sigma^2_{Cr,k} = \frac{1}{m} \sum_{i=1}^{m} \left(j_i - \mu_{Cr} \right)^2 \tag{4.12}$$

where m refers to the number of Cb and Cr samples in SBs k of the Cb and Cr CBs, respectively (see Figure 4.2). Variables v_i and j_i correspond to the i^{th} Cb and Cr samples in SBs k, respectively. Variables μ_{Cb} and μ_{Cr} denote the mean sample values in SBs k, respectively.

Table 4.1: Tabulated bitrate reduction percentages attained for the proposed CCCPQ technique compared with AdaptiveQP. The bitrate reductions are averaged over five QP data points (initial QPs 17, 22, 27, 37 and 37). The AI results are shown on the left; the RA results are shown on the right. The text in red indicates bitrate inflations.

CCCPQ Versus AdaptiveQP (YCbCr 4:2:0) – AI		**CCCPQ Versus AdaptiveQP (YCbCr 4:2:0) – RA**	
Sequence	Bitrate (%)	**Sequence**	Bitrate (%)
BirdsInCage	**−8.5**	BirdsInCage	**−0.1**
DuckAndLegs	**−3.7**	DuckAndLegs	**−1.1**
Kimono	**−17.4**	Kimono	**−7.6**
OldTownCross	**−15.2**	OldTownCross	**−8.1**
ParkScene	**−7.6**	ParkScene	**−3.6**
Traffic	**−13.0**	Traffic	**−6.6**
CCCPQ Versus AdaptiveQP (YCbCr 4:2:2) – AI		**CCCPQ Versus AdaptiveQP (YCbCr 4:2:2) – RA**	
Sequence	Bitrate (%)	**Sequence**	Bitrate (%)
BirdsInCage	**−10.3**	BirdsInCage	0.5
DuckAndLegs	**−5.2**	DuckAndLegs	**−3.0**
Kimono	**−19.8**	Kimono	**−11.5**
OldTownCross	**−18.2**	OldTownCross	**−10.4**
ParkScene	**−13.5**	ParkScene	**−8.2**
Traffic	**−13.2**	Traffic	**−6.4**
CCCPQ Versus AdaptiveQP (YCbCr 4:4:4) – AI		**CCCPQ Versus AdaptiveQP (YCbCr 4:4:4) – RA**	
Sequence	Bitrate (%)	**Sequence**	Bitrate (%)
BirdsInCage	**−15.5**	BirdsInCage	**−3.3**
DuckAndLegs	**−7.9**	DuckAndLegs	**−4.1**
Kimono	**−22.5**	Kimono	**−10.3**
OldTownCross	**−21.0**	OldTownCross	**−16.2**
ParkScene	**−16.8**	ParkScene	**−5.7**
Traffic	**−15.7**	Traffic	**−7.5**

4.2.1 CCCPQ: Experimental Evaluations, Results and Discussion

The same experimental evaluation procedure is employed in all three contribution chapters of this PhD thesis. Please refer to section 1.2 in Chapter 1 for a detailed overview of the conditions employed in the objective and subjective evaluations.

CCCPQ is evaluated and compared with the AdaptiveQP technique, which is the default perceptual quantisation technique in the HEVC standard. It is important to affirm that AdaptiveQP, by definition, outperforms URQ in terms of achieving bitrate reductions without affecting perceptual quality [39]-[41]. Therefore, comparing CCCPQ with AdaptiveQP also implies comparing it with URQ.

Table 4.2: The SSIM results for the 'CCCPQ versus the raw data' in comparison with 'AdaptiveQP versus the raw data' (initial QPs 17, 22, 27, 32 and 37) using the AI encoding configuration. Green text indicates superior results.

Mean SSIM Values (Per Sequence, Per QP): CCCPQ Versus AdaptiveQP (YCbCr 4:2:0) – All Intra

Sequence	CCCPQ					AdaptiveQP				
	QP 17	QP 22	QP 27	QP 32	QP 37	QP 17	QP 22	QP 27	QP 32	QP 37
BirdsInCage	0.9940	0.9914	0.9885	0.9856	0.9822	0.9945	0.9913	0.9884	0.9857	0.9820
DuckAndLegs	0.9870	0.9684	0.9292	0.8960	0.8600	0.9892	0.9724	0.9371	0.8983	0.8615
Kimono	0.9678	0.9441	0.9271	0.9105	0.8909	0.9739	0.9504	0.9288	0.9116	0.8917
OldTownCross	0.9660	0.9219	0.8737	0.8417	0.8057	0.9793	0.9492	0.8970	0.8399	0.7995
ParkScene	0.9757	0.9590	0.9357	0.9024	0.8570	0.9810	0.9625	0.9373	0.9026	0.8561
Traffic	0.9819	0.9650	0.9458	0.9199	0.8841	0.9860	0.9695	0.9520	0.9264	0.8920

Mean SSIM Values (Per Sequence, Per QP): CCCPQ Versus AdaptiveQP (YCbCr 4:2:2) – All Intra

Sequence	CCCPQ					AdaptiveQP				
	QP 17	QP 22	QP 27	QP 32	QP 37	QP 17	QP 22	QP 27	QP 32	QP 37
BirdsInCage	0.9936	0.9903	0.9869	0.9835	0.9789	0.9943	0.9903	0.9867	0.9834	0.9794
DuckAndLegs	0.9881	0.9695	0.9248	0.8714	0.8268	0.9901	0.9733	0.9335	0.8794	0.8283
Kimono	0.9677	0.9368	0.9078	0.8877	0.8630	0.9745	0.9464	0.9139	0.8893	0.8651
OldTownCross	0.9664	0.9142	0.8451	0.8090	0.7719	0.9800	0.9478	0.8875	0.8104	0.7630
ParkScene	0.9675	0.9278	0.8770	0.8380	0.7947	0.9780	0.9459	0.8953	0.8420	0.7965
Traffic	0.9839	0.9674	0.9477	0.9189	0.8761	0.9878	0.9726	0.9541	0.9269	0.8863

Mean SSIM Values (Per Sequence, Per QP): CCCPQ Versus AdaptiveQP (YCbCr 4:4:4) – All Intra

Sequence	CCCPQ					AdaptiveQP				
	QP 17	QP 22	QP 27	QP 32	QP 37	QP 17	QP 22	QP 27	QP 32	QP 37
BirdsInCage	0.9925	0.9870	0.9818	0.9769	0.9730	0.9939	0.9879	0.9816	0.9768	0.9730
DuckAndLegs	0.9918	0.9772	0.9325	0.8655	0.8023	0.9934	0.9805	0.9409	0.8798	0.8122
Kimono	0.9710	0.9376	0.8957	0.8677	0.8457	0.9772	0.9486	0.9082	0.8726	0.8473
OldTownCross	0.9720	0.9220	0.8239	0.7462	0.7092	0.9831	0.9524	0.8866	0.7897	0.7057
ParkScene	0.9715	0.9340	0.8759	0.8260	0.7845	0.9805	0.9513	0.9010	0.8392	0.7864
Traffic	0.9850	0.9691	0.9485	0.9205	0.8815	0.9890	0.9751	0.9559	0.9285	0.8912

The overall bitrate reduction results achieved by CCCPQ are shown in Table 4.1. The most significant bitrate reduction is gained on the Kimono 4:4:4 sequence in the AI tests. That is, over five QP data points, a significant bitrate reduction of 22.5% is accomplished. In the AI QP = 27 test conducted on this sequence, a bitrate reduction of 32.5% is attained compared with AdaptiveQP. The subjective evaluation participants revealed that no perceptual differences can be discerned between the CCCPQ-coded versions and the AdaptiveQP-coded versions of the Kimono 4:4:4 sequences (in all tests for all QPs). Also, compared with the raw data, Pixel-PAQ achieves visually lossless coding in the QP = 22 and QP = 27 (AI and RA) tests.

Table 4.3: The SSIM results for the 'CCCPQ versus the raw data' in comparison with 'AdaptiveQP versus the raw data' (initial QPs 17, 22, 27, 32 and 37) using the RA encoding configuration. Green text indicates superior results.

Mean SSIM Values (Per Sequence, Per QP): CCCPQ Versus AdaptiveQP (YCbCr 4:2:0) – Random Access

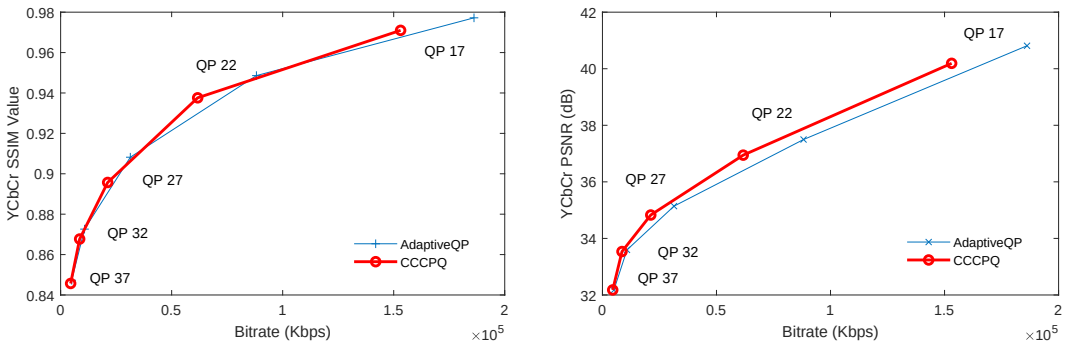
Sequence	CCCPQ					AdaptiveQP				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	**0.9904**	**0.9898**	**0.9885**	**0.9862**	0.9831	**0.9904**	**0.9898**	**0.9885**	**0.9862**	**0.9832**
DuckAndLegs	**0.9476**	**0.9198**	**0.9059**	0.8819	0.8460	0.9391	**0.9198**	**0.9059**	**0.8820**	**0.8475**
Kimono	**0.9319**	**0.9256**	0.9132	0.8965	0.8772	0.9312	0.9252	**0.9137**	**0.8974**	**0.8780**
OldTownCross	**0.9028**	**0.8648**	**0.8569**	**0.8416**	**0.8129**	0.8814	0.8632	0.8540	0.8372	0.8079
ParkScene	**0.9623**	**0.9503**	0.9283	0.8961	0.8526	0.9618	0.9502	**0.9295**	**0.8967**	**0.8530**
Traffic	0.9745	0.9569	0.9417	0.9187	0.8869	**0.9755**	**0.9586**	**0.9457**	**0.9241**	**0.8931**

Mean SSIM Values (Per Sequence, Per QP): CCCPQ Versus AdaptiveQP (YCbCr 4:2:2) – Random Access

Sequence	CCCPQ					AdaptiveQP				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	**0.9882**	0.9869	**0.9859**	0.9836	0.9797	0.9881	**0.9870**	**0.9859**	**0.9837**	**0.9800**
DuckAndLegs	**0.9526**	**0.8962**	**0.8821**	0.8557	0.8149	0.9470	0.8960	0.8820	**0.8561**	**0.8158**
Kimono	**0.9081**	**0.9018**	**0.8918**	0.8749	0.8527	0.9073	0.9017	0.8914	**0.8756**	**0.8543**
OldTownCross	**0.8784**	**0.8228**	**0.8165**	**0.8023**	**0.7720**	0.8374	0.8218	0.8144	0.7971	0.7661
ParkScene	**0.8945**	0.8828	**0.8653**	0.8368	0.7986	0.8934	**0.8829**	0.8649	**0.8376**	**0.7991**
Traffic	0.9741	0.9573	0.9435	0.9202	0.8837	**0.9754**	**0.9592**	**0.9470**	**0.9258**	**0.8911**

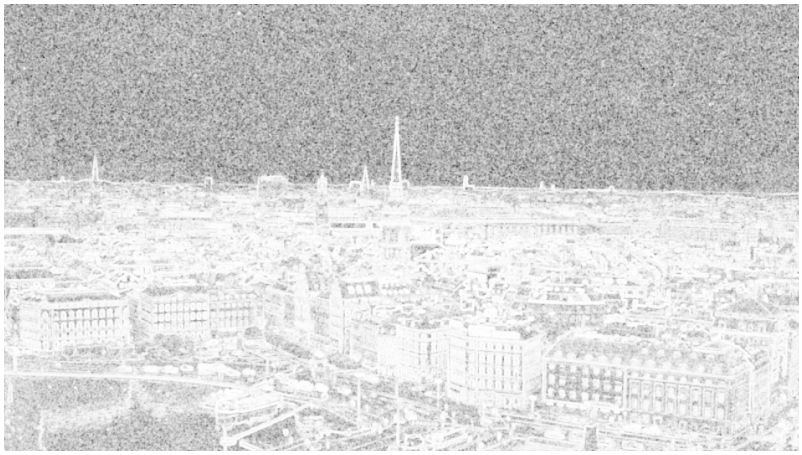
Mean SSIM Values (Per Sequence, Per QP): CCCPQ Versus AdaptiveQP (YCbCr 4:4:4) – Random Access

Sequence	CCCPQ					AdaptiveQP				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	**0.9833**	**0.9784**	0.9775	**0.9759**	0.9729	0.9827	0.9783	**0.9776**	**0.9759**	**0.9730**
DuckAndLegs	**0.9682**	**0.9044**	**0.8526**	**0.8271**	0.7855	0.9656	0.8882	0.8517	0.8268	**0.7866**
Kimono	**0.9133**	**0.8756**	**0.8674**	**0.8537**	0.8341	0.8973	**0.8756**	0.8673	**0.8537**	**0.8347**
OldTownCross	**0.8940**	**0.7493**	**0.7481**	**0.7360**	**0.7109**	0.8715	0.7475	0.7450	0.7304	0.7038
ParkScene	**0.9145**	0.8642	0.8491	0.8238	0.7872	0.8855	**0.8651**	**0.8495**	**0.8240**	**0.7886**
Traffic	0.9727	0.9550	0.9406	0.9188	0.8860	**0.9740**	**0.9573**	**0.9442**	**0.9240**	**0.8926**



(a) (b)

Figure 4.3: Two plots which highlight the bitrate reductions attained by CCCPQ compared with AdaptiveQP. The subfigures show the bitrate reductions achieved by CCCPQ on the following sequences. Subfigure (a): Kimono 4:4:4 (AI - SSIM). Subfigure (b): Kimono 4:4:4 (AI - PSNR).



(a) Luma Channel

(b) Chroma Cb Channel

(c) Chroma Cr Channel

Figure 4.4: The SSIM Index Map (structural reconstruction errors) of a CCCPQ-coded intra-frame (AI QP = 22 test) versus the raw data (OldTownCross 4:4:4 sequence). In subfigures (a), (b) and (c), respectively, the luma (Y), chroma Cb and chroma Cr structural reconstruction errors are shown separately. Note that these reconstruction errors in the CCCPQ-coded compressed sequence are imperceptible to the HVS according to the subjective evaluations (compare the subfigures in Figure 4.5).

(a) CCCPQ-Coded Intra-Frame (AI = QP 22): YCbCr PSNR = 34.6762 and YCbCr SSIM = 0.9220

(b) Raw Data

Figure 4.5: A frame from the OldTownCross 4:4:4 sequence. Subfigure (a) is a CCCPQ-coded intra-frame from this sequence (AI QP = 22 test). Subfigure (b) is the corresponding raw data. Note that the CCCPQ-coded sequence in (a) is perceptually indistinguishable from the raw data in (b) according to the subjective evaluations.

Table 4.4: The PSNR results for the 'CCCPQ versus the raw data' compared with 'AdaptiveQP versus the raw data' (initial QPs 17, 22, 27, 32 and 37) using the AI encoding configuration. Green text indicates superior results.

Mean PSNR (dB) Per Sequence, Per QP: CCCPQ Versus AdaptiveQP (YCbCr 4:2:0) – All Intra

Sequence	CCCPQ					AdaptiveQP				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	41.6270	**39.3906**	**37.5636**	**36.0744**	**34.6347**	**41.8157**	39.2369	37.3972	35.9675	34.5373
DuckAndLegs	39.5321	35.5828	31.9098	**29.7334**	**27.9149**	**39.7520**	**35.7159**	**32.0010**	29.6863	27.8637
Kimono	40.2888	37.8771	36.2440	**34.7342**	**33.1532**	**40.7871**	**38.1203**	**36.2740**	34.7106	33.1212
OldTownCross	39.2658	35.5752	33.1489	**31.5344**	**29.9519**	**40.0628**	**36.2392**	**33.2615**	31.1866	29.5607
ParkScene	40.6118	**37.4512**	**34.5007**	**31.9460**	**29.6601**	**40.7720**	37.3195	34.3177	31.7380	29.4794
Traffic	40.8852	37.6162	35.2187	32.9715	30.8219	**41.3648**	**37.8162**	**35.4051**	**33.0592**	**30.8987**

Mean PSNR (dB) Per Sequence, Per QP: CCCPQ Versus AdaptiveQP (YCbCr 4:2:2) – All Intra

Sequence	CCCPQ					AdaptiveQP				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	41.3116	**38.9338**	**36.9749**	**35.4889**	**33.9824**	**41.6078**	38.7626	36.8004	35.3596	33.9172
DuckAndLegs	39.4774	35.3838	31.4034	28.8754	**27.0114**	**39.7308**	**35.4984**	**31.5177**	**28.8786**	26.9435
Kimono	40.0865	37.2604	35.5054	**34.0871**	**32.4569**	**40.6866**	**37.6684**	**35.6166**	34.0790	32.4313
OldTownCross	39.0418	35.0172	32.3326	**30.8515**	**29.3445**	**39.9615**	**35.8814**	**32.7637**	30.5620	28.9740
ParkScene	39.5213	35.9227	33.1352	31.0041	29.1033	**40.2072**	**36.3055**	**33.1986**	30.8749	28.9633
Traffic	41.3214	37.7927	35.2788	32.7979	30.3881	**41.8157**	**38.0444**	**35.4293**	**32.9055**	**30.4697**

Mean PSNR (dB) Per Sequence, Per QP: CCCPQ Versus AdaptiveQP (YCbCr 4:4:4) – All Intra

Sequence	CCCPQ					AdaptiveQP				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	40.7603	37.8186	**35.6896**	**34.1277**	**33.0549**	**41.3324**	**37.9698**	35.5820	34.0777	33.0049
DuckAndLegs	40.4708	36.1184	31.2134	28.3183	26.4805	**40.8795**	**36.2983**	**31.4508**	**28.5131**	**26.5169**
Kimono	40.1907	36.9447	34.8281	33.5360	**32.1785**	**40.8125**	**37.4973**	**35.1411**	**33.5945**	32.1699
OldTownCross	39.0588	34.6762	31.2903	29.6014	**28.4592**	**40.0621**	**35.6320**	**32.1010**	**29.7578**	28.1781
ParkScene	39.7764	36.0497	33.1394	**31.0266**	**29.1732**	**40.5012**	**36.5698**	**33.4090**	31.0139	29.0532
Traffic	41.6707	38.4761	35.5334	32.9007	30.4820	**42.2675**	**38.8464**	**35.7583**	**33.0303**	**30.5622**

In most cases, AdaptiveQP attains higher PSNR and SSIM values compared with CCCPQ, particularly in the AI tests. This is by virtue of the increased levels of perceptual quantisation applied to high variance regions of the video data (see Table 4.2 to Table 4.5). However, CCCPQ achieves marginal reconstruction quality improvements in most of the RA tests. Recall that the RA configuration employs bidirectional inter prediction. Therefore, due to the GOP-based nature of inter prediction, the frame-level luma and chroma QP offsets associated with the GOP structure can potentially affect QP selection at the $2N \times 2N$ CU level when computing the spatial variances of the raw video data. Evidently, this does not affect the overall bitrate reductions attained by CCCPQ.

Table 4.5: The PSNR results for the 'CCCPQ versus the raw data' compared with 'AdaptiveQP versus the raw data' (initial QPs 17, 22, 27, 32 and 37) using the RA encoding configuration. Green text indicates superior results.

Mean PSNR (dB) Per Sequence, Per QP: CCCPQ Versus AdaptiveQP (YCbCr 4:2:0) – Random Access

Sequence	CCCPQ					AdaptiveQP				
	QP 17	QP 22	QP 27	QP 32	QP 37	QP 17	QP 22	QP 27	QP 32	QP 37
BirdsInCage	**38.6438**	**38.2302**	**37.4348**	**36.2584**	**34.8998**	38.5969	38.1917	37.3700	36.1888	34.8880
DuckAndLegs	**33.4449**	**31.4966**	**30.2525**	**28.7110**	**27.0648**	33.1007	31.4451	30.2041	28.6417	27.0503
Kimono	**37.1740**	**36.2955**	**34.8949**	33.2905	**31.8207**	37.1056	36.2425	34.8777	**33.3143**	31.7884
OldTownCross	**34.5946**	**32.9144**	**32.4098**	**31.5443**	**30.2423**	33.8699	32.8271	32.2600	31.3035	29.9849
ParkScene	**37.9692**	**35.9822**	**33.7099**	**31.5642**	**29.5759**	37.7894	35.8050	33.6040	31.4737	29.4641
Traffic	**38.5707**	36.2078	34.4272	32.5880	30.7535	38.5707	**36.2246**	**34.5417**	**32.6907**	**30.8408**

Mean PSNR (dB) Per Sequence, Per QP: CCCPQ Versus AdaptiveQP (YCbCr 4:2:2) – Random Access

Sequence	CCCPQ					AdaptiveQP				
	QP 17	QP 22	QP 27	QP 32	QP 37	QP 17	QP 22	QP 27	QP 32	QP 37
BirdsInCage	**37.7817**	**37.0259**	**36.5453**	**35.5843**	**34.2365**	37.7182	37.0119	36.5037	35.5372	34.2150
DuckAndLegs	**33.2208**	**30.3237**	**29.3665**	**27.9316**	**26.2980**	32.8991	30.2871	29.3192	27.8967	26.2511
Kimono	**35.8379**	**35.2210**	**34.1771**	32.7507	31.2806	35.7738	35.1864	34.1301	**32.7657**	**31.2820**
OldTownCross	**33.4955**	**31.9207**	**31.5955**	**30.8824**	**29.6354**	32.5644	31.8778	31.4593	30.6358	29.3742
ParkScene	**34.5726**	**33.5462**	**32.2378**	**30.6668**	**29.0242**	34.4633	33.4402	32.1086	30.5555	28.9179
Traffic	38.4694	36.0255	34.3894	32.5485	30.5302	**38.4959**	**36.0539**	**34.4692**	**32.6626**	**30.6154**

Mean PSNR (dB) Per Sequence, Per QP: CCCPQ Versus AdaptiveQP (YCbCr 4:4:4) – Random Access

Sequence	CCCPQ					AdaptiveQP				
	QP 17	QP 22	QP 27	QP 32	QP 37	QP 17	QP 22	QP 27	QP 32	QP 37
BirdsInCage	**36.6496**	**34.7491**	34.5005	**34.0177**	**33.1848**	36.4722	34.7233	**34.5008**	34.0001	33.1721
DuckAndLegs	**34.1487**	**29.5518**	**28.0018**	**26.8944**	**25.4593**	33.8200	29.2282	27.9528	26.8411	25.4314
Kimono	**35.6900**	**34.1614**	**33.4209**	**32.2993**	**31.0126**	35.1569	34.1381	33.3936	32.2908	31.0046
OldTownCross	**33.4595**	**30.1233**	**30.0231**	**29.5890**	**28.7174**	32.6894	30.0705	29.9561	29.4221	28.4866
ParkScene	**35.0219**	**33.0549**	**32.0072**	**30.6318**	**29.0461**	34.2332	33.0092	31.9277	30.5388	28.9655
Traffic	38.3777	36.0263	34.1664	32.3883	30.5110	**38.3829**	**36.0421**	**34.2489**	**32.4925**	**30.5736**

The subfigures in Figure 4.4 highlight the structural reconstruction errors for each colour channel in the OldTownCross 4:4:4 sequence. Identical in context to the examples shown in Figure 3.7 of Chapter 3 in this thesis, the structural reconstruction errors for the coded version of the OldTownCross 4:4:4 sequence are concentrated mostly in the high variance regions in each channel of the YCbCr video data. Recall that CCCPQ exploits the variance-based spatial masking phenomenon of the HVS. Therefore, the reconstruction errors incurred by the proposed CCCPQ technique (Figure 4.4) are imperceptible to the HVS; compare the subfigures in Figures 4.5.

Table 4.6: The Mean Opinion Score (MOS) results, rounded to the nearest integer, of four participants in the subjective evaluations for CCCPQ versus AdaptiveQP.

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – CCCPQ versus AdaptiveQP

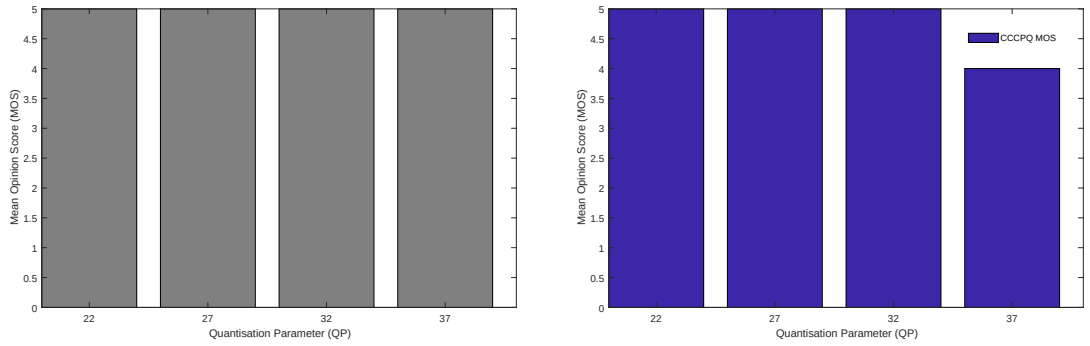
Sequence	YCbCr 4:2:0 All Intra				YCbCr 4:2:0 Random Access			
	QP 22	**QP 27**	**QP 32**	**QP 37**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	5	5	5	5	5	5	5	5
DuckAndLegs	5	5	5	5	5	5	5	5
Kimono	5	5	5	5	5	5	5	5
OldTownCross	5	5	5	5	5	5	5	5
ParkScene	5	5	5	5	5	5	5	5
Traffic	5	5	5	5	5	5	5	5

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – CCCPQ versus AdaptiveQP

Sequence	YCbCr 4:2:2 All Intra				YCbCr 4:2:2 Random Access			
	QP 22	**QP 27**	**QP 32**	**QP 37**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	5	5	5	5	5	5	5	5
DuckAndLegs	5	5	5	5	5	5	5	5
Kimono	5	5	5	4	5	5	5	4
OldTownCross	5	5	5	5	5	5	5	5
ParkScene	5	5	5	5	5	5	5	5
Traffic	5	5	5	5	5	5	5	5

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – CCCPQ versus AdaptiveQP

Sequence	YCbCr 4:4:4 All Intra				YCbCr 4:4:4 Random Access			
	QP 22	**QP 27**	**QP 32**	**QP 37**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	5	5	5	4	5	5	5	5
DuckAndLegs	5	5	5	5	5	5	5	5
Kimono	5	5	5	4	5	5	5	5
OldTownCross	5	5	5	5	5	5	5	4
ParkScene	5	5	5	5	5	5	5	5
Traffic	5	5	5	5	5	5	5	5



(a) (b)

Figure 4.6: Two Mean Opinion Score (MOS) bar graphs. Subfigure (a) shows the MOS for CCCPQ versus AdaptiveQP on the vast majority of sequences using the AI and RA configurations. Subfigure (b) shows the MOS for CCCPQ versus the raw video data on the Traffic 4:4:4 10-bit sequence.

Table 4.7: The Mean Opinion Score (MOS) results, rounded to the nearest integer, of four participants in the subjective evaluations for CCCPQ versus the raw data.

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – CCCPQ versus Raw Data

Sequence	YCbCr 4:2:0 All Intra				YCbCr 4:2:0 Random Access			
	QP 22	QP 27	QP 32	QP 37	QP 22	QP 27	QP 32	QP 37
BirdsInCage	5	4	3	2	5	5	5	4
DuckAndLegs	5	5	4	3	5	5	4	4
Kimono	5	5	4	3	5	5	4	3
OldTownCross	5	4	3	2	5	4	4	3
ParkScene	5	4	3	2	5	4	3	3
Traffic	5	5	4	3	5	5	4	4

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – CCCPQ versus Raw Data

Sequence	YCbCr 4:2:2 All Intra				YCbCr 4:2:2 Random Access			
	QP 22	QP 27	QP 32	QP 37	QP 22	QP 27	QP 32	QP 37
BirdsInCage	5	4	3	2	5	5	4	4
DuckAndLegs	5	4	4	3	5	5	4	3
Kimono	5	4	3	2	5	5	4	2
OldTownCross	5	4	3	2	5	5	5	4
ParkScene	5	4	3	2	5	5	4	3
Traffic	5	5	4	3	5	5	5	4

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – CCCPQ versus Raw Data

Sequence	YCbCr 4:4:4 All Intra				YCbCr 4:4:4 Random Access			
	QP 22	QP 27	QP 32	QP 37	QP 22	QP 27	QP 32	QP 37
BirdsInCage	5	5	4	3	5	5	4	4
DuckAndLegs	5	5	4	3	5	5	4	4
Kimono	5	5	4	3	5	5	4	3
OldTownCross	5	5	3	2	5	5	4	3
ParkScene	5	5	3	2	5	5	4	3
Traffic	5	5	4	3	5	5	5	4

The participants of the subjective evaluations confirmed that the CCCPQ-coded sequences and the AdaptiveQP-coded sequences are perceived as perceptually identical in the vast majority of cases (see the MOS scores in Table 4.6). In contrast with the TCPQ subjective evaluation results included in Chapter 3 in this thesis, the subjective evaluation results in Table 4.6 correlate closely with the SSIM and PSNR objective visual quality results conducted for CCCPQ versus AdaptiveQP. Recall that, in many cases, the CCCPQ-coded sequences are marginally superior in terms of mathematical reconstruction quality compared with AdaptiveQP-coded sequences, as quantified by SSIM and PSNR.

The MOS scores in Table 4.7 indicate that, in many cases, the CCCPQ-coded sequences are perceptually indistinguishable from the raw video data; therefore, visually lossless coding is triumphantly accomplished. For instance, in the following sequences: BirdsInCage 4:2:0, OldTownCross 4:2:2, Traffic 4:2:2 and Traffic 4:4:4 in the RA QP = 22, QP = 27 and QP = 32, the participants perceived the CCCPQ-coded sequences to be perceptually identical to the corresponding raw video data. This is significant due to the fact that visually lossless coding is achieved at a high initial QP (i.e., QP = 32). It is reasonable to infer that CCCPQ attains visually lossless coding at high initial QP values because of the CU-level QP adjustment approach. That is, in comparison with AdaptiveQP, CCCPQ provides a more wholesome computation of spatial activity in the $2N\times2N$ CU; this drives the efficacy of the cross-colour channel dependency for QP selection mechanism. Consequently, a more appropriate CU-level QP can be selected, thereby achieving a useful balance between bitrate reductions and HVS-based perceptual reconstruction quality.

As is the case with the proposed TCPQ technique presented in Chapter 3 of this thesis, the CCCPQ bitrate reduction evaluations results provide empirical evidence that CCCPQ is most effective when applied to 4:4:4 sequences. It is less effective when applied to 4:2:2 sequences and less effective still when applied to 4:2:0 sequences. It is reasonable to infer that CCCPQ is less effective on chroma subsampled versions of each sequence because the Cb and Cr spatial downsampling data results in Cb CB and Cr CB size reductions in the corresponding $2N\times2N$ CU. In other words, a smaller chroma CB equates to inferior spatial activity computations with respect to the constituent Cb CB and the Cr CB SBs, thus resulting in the selection of a less appropriate $2N\times2N$ CU level QP.

The MOS scores in the RA tests in the ‘CCCPQ versus the raw data’ experiments are, in general, significantly superior compared with the corresponding AI tests; this is especially true for high initial QP values. Recall that this also proved to be the case for the proposed TCPQ technique in Chapter 3 of this thesis. To reiterate from Chapter 3, this is because GOP-based inter coding includes the signalling of important motion data in the bitstream [10, 11], whereas All Intra coding does not account for motion data or spatiotemporal redundancies [9].

4.3 Proposed CB-Level Perceptual Quantisation Method for 4:4:4 Data

CBPQ [23] is a CB-level perceptual quantization contribution, which, in contrast to both CCCPQ and AdaptiveQP, is designed to perceptually adjust the QP at the CB level. In other words, CBPQ separately adjusts the QP for the Y CB, the Cb CB and the Cr CB based on the sample variances in each CB. That is, while CCCPQ computes a single QP per $2N \times 2N$ CU level, CBPQ computes three QPs per $2N \times 2N$ CU (i.e., one QP for the Y CB, one QP for the Cb CB and one QP for the Cr CB). CBPQ is tailored for high bit depth 4:4:4 video data. This is because the variance of Y, Cb and Cr samples in high bit depth 4:4:4 video data are considerably different from the corresponding variances in 8-bit subsampled 4:2:0 video data. Compared with 8-bit subsampled 4:2:0 video data, 10-bit 4:4:4 video data contains a much greater number of colour variations in each pixel. Therefore, the higher the bit depth of the video data, in addition to an absence of chroma subsampling, the more potential there is for the data in the chroma Cb and Cr channels to comprise higher variances.

Similar to CCCPQ, CBPQ is primarily designed to decrease coding bitrates without incurring a perceptually discernible decrease in reconstruction quality; this is achieved by increasing the levels of perceptual quantisation applied to luma and chroma data. To reiterate, however, it is tailored specifically for 4:4:4 video data. In the case of the AdaptiveQP method discussed in section 4.1.2, this technique applies an inappropriate CU-level QP adjustment when coding 4:4:4 video data. Recall that the primary reason for this is as follows: Adaptive QP computes the variance of raw luma samples in a luma CB only when selecting a CU-level QP. This means that the entire CU, which includes chroma Cb and Cr CBs, is perceptually coded based on the variance of luma data only, which is a significant problem. This is particularly pertinent to high bit depth 4:4:4 video data. The chroma Cb and Cr channels in, for example, 10-bit 4:4:4 video sequences may contain data with significantly higher variances compared with the variances of the data in the corresponding luma channel. This is by virtue of the nature of 10-bit 4:4:4 video data: i.e., 30-bits per pixel (a maximum of 1024^3 unique colours per pixel) in addition to an absence of chroma subsampling. Therefore, for the variance-based perceptual quantisation of high bit depth 4:4:4 video data, separately adjusting the QP for the Y CB, the Cb CB and the Cr CB is desirable.

Due to the nature of high bit depth YCbCr 4:4:4 video data, particularly in relation to a higher probability of higher variance in the chroma channels, separately computing the variances of the raw sample data in the Y CB, the Cb CB and the Cr CB equates to a more refined approach in terms of perceptually increasing the QP according to spatial activity. It is important to note that, in the standardised Format Range Extensions of HEVC, JCT-VC provides the flexibility for signalling chroma QP offsets at the CB level in the PPS [86] in HEVC HM [29, 30]. CBPQ takes advantage of this flexibility, whereas CCCPQ and AdaptiveQP do not. Furthermore, due to the variance-based nature of CBPQ, the proposed method is, by definition, compatible with standardised ITU-R BT.2020 Ultra HD data [87] and ITU-R BT.2100 HDR and Wide Colour Gamut (WCG) data [88].

To reiterate, CBPQ is designed primarily to exploit the aforementioned spatial masking phenomenon of the HVS — for high spatial activity regions in a picture — by increasing quantisation levels for high variance luma and chroma CBs. Unlike CCCPQ, however, it is designed to work with high bit depth 4:4:4 video data. The increased levels of quantisation noise induced by CBPQ gives rise to useful bitrate reductions without incurring a noticeable decrease in perceptual quality. Note that, similar to CCCPQ, CBPQ can also decrease the QP if the luma and chroma CBs are calculated as having low variance values. Therefore, when a low spatial activity CB is detected, the CB-level QP is decreased, thus potentially leading to reconstruction quality improvements.

Recall that AdaptiveQP accounts for the spatial activity in a luma CB; therefore, CBPQ utilises this functionality. Consequently, the operations described in equations (4.1)-(4.6) are employed in CBPQ. To this end, the chroma Cb and Cr CB-level perceptual QPs, denoted as Q_{Cb} and Q_{Cr}, are computed in (4.13) and (4.14), respectively:

$$Q_{Cb} = QP + \left[6 \times \log_2 \left(B \right) \right] \tag{4.13}$$

$$Q_{Cr} = QP + \left[6 \times \log_2 \left(R \right) \right] \tag{4.14}$$

where B and R refer to the normalised spatial activity of chroma Cb and Cr CBs, respectively. B and R are computed in (4.15) and (4.16), respectively:

$$B = \frac{f \cdot g + t_{Cb}}{g + f \cdot t_{Cb}} \tag{4.15}$$

$$R = \frac{f \cdot x + t_{Cr}}{x + f \cdot t_{Cr}} \tag{4.16}$$

where g and x denote the non-normalized spatial activity of chroma Cb and Cr CBs, respectively. Variables t_{Cb} and t_{Cr} refer to the mean spatial activity of all $2N \times 2N$ chroma Cb and Cr CBs belonging to the current picture, respectively. Variables g, t_{Cb}, x and t_{Cr} are computed in (4.17)-(4.20), respectively:

$$g = 1 + \min\left(\sigma^2{}_{Cb,d}\right), \quad \text{where } d = 1,....,4 \tag{4.17}$$

$$t_{Cb} = \frac{1}{C_{Cb}} \sum_{n=1}^{C_{Cb}} g_n \tag{4.18}$$

$$x = 1 + \min\left(\sigma^2{}_{Cb,z}\right), \quad \text{where } z = 1,....,4 \tag{4.19}$$

$$t_{Cr} = \frac{1}{C_{Cr}} \sum_{n=1}^{C_{Cr}} x_n \tag{4.20}$$

where $\sigma^2{}_{Cb,d}$ and $\sigma^2{}_{Cr,z}$ refer to the variances of sample values in SB d of the chroma Cb CB and SB z of the chroma Cr CB, respectively. Variables C_{Cb} and C_{Cr} denote the number of $2N \times 2N$ chroma Cb CBs and Cr CBs in the current picture, respectively. Variables $\sigma^2{}_{Cb,d}$ and $\sigma^2{}_{Cr,z}$ are quantified in (4.21) and (4.22), respectively:

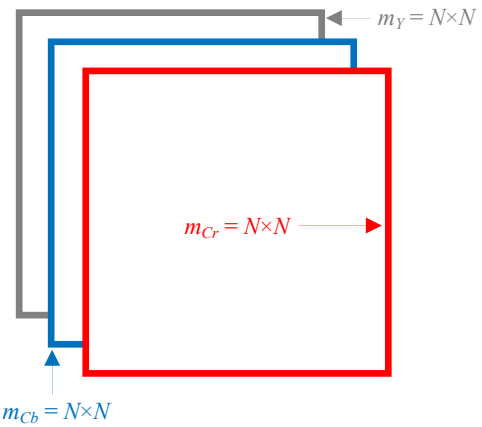


Figure 4.7: The size of the constituent Y (Gray), Cb (Blue) and Cr (Red) CB SBs, denoted as *k*, *d* and *z*, respectively, in CBPQ for 4:4:4 video data.

$$\sigma^2_{Cb,d} = \frac{1}{m_{Cb}} \sum_{n=1}^{m_{Cb}} \left(v_n - \mu_{Cb}\right)^2 \tag{4.21}$$

$$\sigma^2_{Cr,z} = \frac{1}{m_{Cr}} \sum_{n=1}^{m_{Cr}} \left(j_n - \mu_{Cr}\right)^2 \tag{4.22}$$

where m_{Cb} and m_{Cr} denote the number of sample values in SBs *d* and *z*, respectively (see Figure 4.7), where v_n and j_n refer to the n^{th} sample values in SBs *d* and *z*, respectively, and where μ_{Cb} and μ_{Cr} correspond to the mean sample values of SBs *d* and *z*, respectively.

4.3.1 CBPQ: Experimental Evaluations, Results and Discussion

The same experimental evaluation procedure is employed in all three contribution chapters of this PhD thesis. Please refer to section 1.2 in Chapter 1 for a detailed overview of the conditions employed in the objective and subjective evaluations.

CBPQ is evaluated and compared with the AdaptiveQP technique in [39]-[41], which has been previously proposed for the HEVC standard. It is important to affirm that AdaptiveQP, by definition, outperforms URQ in terms of bitrate reductions achieved without affecting perceptual quality [39]-[41]. Therefore, comparing CBPQ with AdaptiveQP also implies comparing it with URQ.

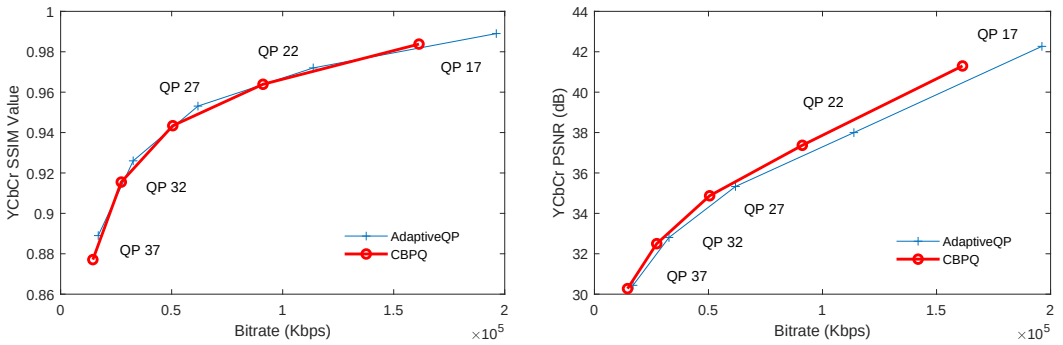
Table 4.8: The overall bitrate reductions attained for the proposed CBPQ technique compared with AdaptiveQP. The bitrates, in Kbps, are averaged over five QP data points (initial QPs 17, 22, 27, 32 and 37). The AI results are shown on the left; the RA results are shown on the right. Green text indicates superior results.

Mean Bitrate (Kbps) – AI (YCbCr 4:4:4)			**Mean Bitrate (Kbps) – RA (YCbCr 4:4:4)**		
Sequence	CBPQ	AdaptiveQP	**Sequence**	CBPQ	AdaptiveQP
BirdsInCage	**100,046**	126,334	BirdsInCage	**22,949**	26,050
DuckAndLegs	**172,355**	191,926	DuckAndLegs	**62,348**	67,527
Kimono	**48,940**	65,573	Kimono	**14,182**	16,705
OldTownCross	**220,703**	283,886	OldTownCross	**75,875**	90,329
ParkScene	**84,320**	103,141	ParkScene	**17,389**	19,228
Traffic	**68,969**	84,278	Traffic	**10,724**	11,806

Table 4.9: The bitrate reductions (%), per sequence averaged over five initial QPs (i.e., QP 17, 22, 27, 32 and 37), attained by the proposed CBPQ technique compared with AdaptiveQP. In addition, the reconstruction quality results (per channel), as quantified by SSIM (%), are tabulated. Green text indicates superior results.

Overall Bitrate (%) Per Sequence and SSIM (%) Per Channel: CBPQ Versus AdaptiveQP (YCbCr 4:4:4)

Sequence	All Intra				Random Access			
	Bitrate	**Y SSIM**	**Cb SSIM**	**Cr SSIM**	**Bitrate**	**Y SSIM**	**Cb SSIM**	**Cr SSIM**
BirdsInCage	**−20.8**	−0.2	−0.4	−0.1	**−11.9**	0.0	0.1	0.0
DuckAndLegs	**−10.2**	−0.6	−1.8	−0.8	**−7.7**	0.0	0.9	−0.1
Kimono	**−25.4**	−0.2	−0.8	−0.2	**−15.1**	0.0	0.5	0.0
OldTownCross	**−22.3**	−0.9	−2.8	−1.2	**−16.0**	0.1	0.2	0.0
ParkScene	**−18.2**	−0.5	−1.4	−0.4	**−9.6**	−0.1	0.8	0.0
Traffic	**−18.2**	−0.7	−0.9	−0.2	**−9.2**	−0.5	−0.2	−0.1



(a) (b)

Figure 4.8: Two plots which highlight the bitrate reductions attained by CBPQ compared with AdaptiveQP. The subfigure show the bitrate reductions achieved by CBPQ on the following sequences. Subfigure (a): Traffic 4:4:4 (AI - SSIM). Subfigure (b): Traffic 4:4:4 (AI - PSNR).

Table 4.10: The SSIM results for the 'CBPQ versus the raw data' in comparison with 'AdaptiveQP versus the raw data' (initial QPs 17, 22, 27, 32 and 37) using the AI and RA configurations. Green text indicates superior results.

Mean SSIM Values (Per Sequence, Per QP): CBPQ Versus AdaptiveQP (YCbCr 4:4:4) – All Intra

Sequence	CBPQ					AdaptiveQP				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	0.9923	0.9864	0.9810	0.9765	0.9728	**0.9939**	**0.9879**	**0.9816**	**0.9768**	**0.9730**
DuckAndLegs	0.9912	0.9749	0.9288	0.8577	0.8010	**0.9934**	**0.9805**	**0.9409**	**0.8798**	**0.8122**
Kimono	0.9700	0.9342	0.8927	0.8671	0.8458	**0.9772**	**0.9486**	**0.9082**	**0.8726**	**0.8473**
OldTownCross	0.9738	0.9250	0.8289	0.7452	0.7056	**0.9831**	**0.9524**	**0.8866**	**0.7897**	**0.7057**
ParkScene	0.9699	0.9281	0.8689	0.8239	0.7835	**0.9805**	**0.9513**	**0.9010**	**0.8392**	**0.7864**
Traffic	0.9838	0.9639	0.9434	0.9155	0.8771	**0.9890**	**0.9720**	**0.9531**	**0.9260**	**0.8890**

Mean SSIM Values (Per Sequence, Per QP): CBPQ Versus AdaptiveQP (YCbCr 4:4:4) – Random Access

Sequence	CBPQ					AdaptiveQP				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	**0.9832**	**0.9783**	**0.9776**	0.9758	0.9728	0.9827	**0.9783**	**0.9776**	**0.9759**	**0.9730**
DuckAndLegs	**0.9668**	**0.9039**	**0.8523**	0.8267	0.7848	0.9656	0.8882	0.8517	**0.8268**	**0.7866**
Kimono	**0.9133**	**0.8757**	**0.8680**	**0.8537**	0.8338	0.8973	0.8756	0.8673	**0.8537**	**0.8347**
OldTownCross	**0.9039**	**0.7483**	**0.7468**	**0.7333**	**0.7076**	0.8715	0.7475	0.7450	0.7304	0.7038
ParkScene	**0.9127**	0.8635	0.8484	0.8226	0.7862	0.8855	**0.8651**	**0.8495**	**0.8240**	**0.7886**
Traffic	0.9722	0.9540	0.9392	0.9171	0.8841	**0.9740**	**0.9573**	**0.9442**	**0.9240**	**0.8926**

The bitrate reduction results in Table 4.8 show that the proposed CBPQ technique consistently outperforms CCCPQ (and also AdaptiveQP) when applied to 4:4:4 video data; compare the results in Table 4.8 with those in Table 4.1. Recall that CBPQ separately adjusts the QP at the CB-level according to the variance computations in each CB. Therefore, this gives rise to a more refined approach in terms of applying perceptual quantisation, especially so for high variance chroma Cb and Cr data.

In terms of the SSIM and PSNR results recorded for CBPQ versus AdaptiveQP, there are no major differences in the overall 'per sequence, per QP' values (see Table 4.10 and Table 4.12). However, as shown in Table 4.9 and 4.11, the 'per channel' SSIM and PSNR results (averaged over five QPs; i.e., initial QPs 17, 22, 27, 32 and 37) show that the CBPQ-coded sequences are, overall, inferior to the AdaptiveQP-coded sequences in terms of objective reconstruction quality. More importantly, however, the subjective evaluation results show that there are no perceptually discernible differences between the CBPQ-coded sequences and the AdaptiveQP-coded sequences. In addition, the participants also confirmed that visually lossless coding is accomplished by CBPQ in all of the AI QP = 22 tests and RA QP = 22 tests.

Table 4.11: The bitrate reductions (%), per sequence averaged over five initial QPs (i.e., QP 17, 22, 27, 32 and 37), attained by the proposed CBPQ technique compared with AdaptiveQP. In addition, the reconstruction quality results (per channel), as quantified by PSNR (%), are tabulated. Green text indicates superior results.

Overall Bitrate (%) Per Sequence and PSNR (%) Per Channel: CBPQ Versus AdaptiveQP (YCbCr 4:4:4)

Sequence	All Intra				Random Access			
	Bitrate	**Y PSNR**	**Cb PSNR**	**Cr PSNR**	**Bitrate**	**Y PSNR**	**Cb PSNR**	**Cr PSNR**
BirdsInCage	**−20.8**	−0.2	−0.8	−0.3	**−11.9**	0.0	0.0	0.0
DuckAndLegs	**−10.2**	−0.5	−1.2	−0.8	**−7.7**	0.2	0.0	0.0
Kimono	**−25.4**	−0.8	−1.4	−0.8	**−15.1**	0.0	0.0	−0.1
OldTownCross	**−22.3**	−0.9	−2.1	−1.3	**−16.0**	0.4	−0.4	0.1
ParkScene	**−18.2**	−0.8	−1.3	−0.8	**−9.6**	−0.1	0.0	−0.1
Traffic	**−18.2**	−1.1	−1.5	−1.1	**−9.2**	−0.5	−0.2	−0.2

Table 4.12: The PSNR results for the 'CBPQ versus the raw data' in comparison with 'AdaptiveQP versus the raw data' (initial QPs 17, 22, 27, 32 and 37) using the AI and RA configurations. Green text indicates superior results.

Mean PSNR (dB) Per Sequence, Per QP: CBPQ Versus AdaptiveQP (YCbCr 4:4:4) – All Intra

Sequence	CBPQ					AdaptiveQP				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	40.5628	37.5369	35.4369	34.0196	32.9895	**41.3324**	**37.9698**	**35.5820**	**34.0777**	**33.0049**
DuckAndLegs	40.2146	35.7731	31.0767	28.2018	26.4413	**40.8795**	**36.2983**	**31.4508**	**28.5131**	**26.5169**
Kimono	40.0946	36.7891	34.7642	33.5335	**32.2021**	**40.8125**	**37.4973**	**35.1411**	**33.5945**	32.1699
OldTownCross	39.1589	34.6405	31.1861	29.4800	**28.3486**	**40.0621**	**35.6320**	**32.1010**	**29.7578**	28.1781
ParkScene	39.6489	35.7914	32.9557	30.9214	**29.1320**	**40.5012**	**36.5698**	**33.4090**	**31.0139**	29.0532
Traffic	41.2975	37.3692	34.8709	32.4997	**30.2729**	**42.2675**	**37.9978**	**35.3301**	**32.8053**	30.4383

Mean PSNR (dB) Per Sequence, Per QP: CBPQ Versus AdaptiveQP (YCbCr 4:4:4) – Random Access

Sequence	CBPQ					AdaptiveQP				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	**36.5972**	**34.7253**	**34.5045**	33.9932	33.1445	36.4722	34.7233	34.5008	**34.0001**	**33.1721**
DuckAndLegs	**34.0262**	**29.5089**	**27.9711**	26.8540	**25.4230**	33.8200	29.2282	27.9528	26.8411	**25.4314**
Kimono	**35.6757**	**34.1822**	**33.4316**	**32.3019**	31.0003	35.1569	34.1381	33.3936	32.2908	**31.0046**
OldTownCross	**33.7468**	**30.1016**	**29.9772**	**29.5086**	**28.6292**	32.6894	30.0705	29.9561	29.4221	28.4866
ParkScene	**34.9486**	**33.0174**	**31.9460**	**30.5524**	**28.9891**	34.2332	33.0092	31.9277	30.5388	28.9655
Traffic	38.2919	35.9170	34.0580	32.2884	30.4332	**38.3829**	**36.0421**	**34.2489**	**32.4925**	**30.5736**

In relation to the 'per sequence, per QP' and 'per channel' SSIM and PSNR results, no major differences are observed between CBPQ and AdaptiveQP. Similar to the results recorded in the CCCPQ tests, marginal reconstruction quality improvements are attained by CBPQ in the RA inter-coding tests. Therefore, as previously mentioned, the frame-level luma and chroma QP offsets associated with the RA GOP structure can affect picture-level QP adjustments related to variance computations.

(a) Luma Channel

(b) Chroma Cb Channel

(c) Chroma Cr Channel

Figure 4.9: The SSIM Index Map (structural reconstruction errors) of a CBPQ-coded inter-frame (RA QP = 22 test) versus the raw data (Kimono 4:4:4 sequence). In subfigures (a), (b) and (c), respectively, the luma (Y), chroma Cb and chroma Cr structural reconstruction errors are shown separately. Note that these reconstruction errors in the CBPQ-coded compressed sequence are imperceptible to the HVS according to the subjective evaluations (compare the subfigures in Figure 4.10).

(a) CBPQ-Coded Inter-Frame (RA = QP 22): YCbCr PSNR = 34.1822 and YCbCr SSIM = 0.8757

(b) Raw Data

Figure 4.10: A frame from the Kimono 4:4:4 sequence. Subfigure (a) is a CBPQ-coded inter-frame from this sequence (RA QP = 22 test). Subfigure (b) is the corresponding raw data. Note that the CBPQ-coded sequence in subfigure (a) is perceptually indistinguishable from the raw data in subfigure (b) according to the subjective evaluations.

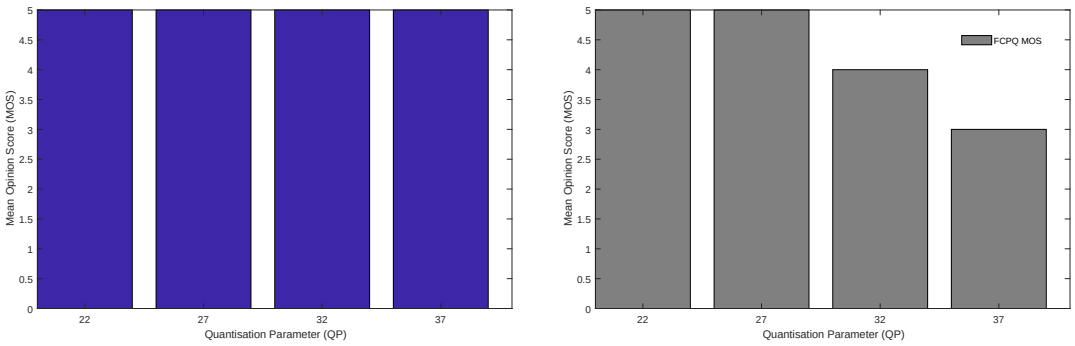
Table 4.13: The Mean Opinion Score (MOS) results, rounded to the nearest integer, of four participants in the subjective evaluations for 'CBPQ versus AdaptiveQP' and also 'CBPQ versus the raw data'. The sequences viewed are coded with initial QPs 17, 22, 27, 32 and 37.

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – CBPQ versus AdaptiveQP

Sequence	YCbCr 4:4:4 All Intra				YCbCr 4:4:4 Random Access			
	QP 22	QP 27	QP 32	QP 37	QP 22	QP 27	QP 32	QP 37
BirdsInCage	5	5	5	4	5	5	5	5
DuckAndLegs	5	5	5	5	5	5	5	5
Kimono	5	5	5	3	5	5	5	4
OldTownCross	5	5	5	5	5	5	5	5
ParkScene	5	5	5	5	5	5	5	5
Traffic	5	5	5	5	5	5	5	5

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – CBPQ versus Raw Data

Sequence	YCbCr 4:4:4 All Intra				YCbCr 4:4:4 Random Access			
	QP 22	QP 27	QP 32	QP 37	QP 22	QP 27	QP 32	QP 37
BirdsInCage	5	4	3	3	5	5	4	4
DuckAndLegs	5	5	4	3	5	5	4	4
Kimono	5	4	3	3	5	5	4	4
OldTownCross	5	5	3	2	5	5	4	4
ParkScene	5	4	3	3	5	5	4	4
Traffic	5	5	4	3	5	5	4	4



(a) (b)

Figure 4.11: Two Mean Opinion Score (MOS) bar graphs. Subfigure (a) shows the MOS for CCCPQ versus AdaptiveQP on the vast majority of sequences using the AI and RA configurations. Subfigure (b) shows the MOS for CCCPQ versus the raw video data on the Traffic 4:4:4 10-bit sequence.

The MOS results in Table 4.13 indicate that the subjective evaluation participants are unable to detect perceptible differences between the CBPQ-coded sequences and the AdaptiveQP-coded sequences in all of the AI and RA QP = 22, QP = 27 and QP = 32 tests. However, in certain AI and RA QP = 37 tests, the quantisation-induced compression artifacts proved to be perceptible.

(a) (b)

Figure 4.12: Two images showing a portion of the DuckAndLegs 4:4:4 sequence. Subfigure (a) shows a portion of the first frame of the DuckAndLegs 4:4:4 raw video data. Subfigure (b) shows a portion of the first intra-coded frame (AI QP = 37 test) of the CBPQ-coded DuckAndLegs 4:4:4 compressed sequence. In subfigure (b), note that the white square highlights a region in which the higher levels of quantisation to the chroma Cb and Cr data is conspicuous when the image is magnified and stationary. That is, in addition to the loss of high frequency detail in the luma data, the colourfulness in the chroma Cb and Cr data is significantly decreased in this region.

With respect to the CBPQ-coded sequences versus the raw video data test results (see Table 4.13), visually lossless coding is accomplished in all of the AI QP = 22 tests in addition to the RA QP = 22 and QP = 27 tests. Due to the fact that CBPQ incurs higher levels of perceptual quantisation to the samples in chroma Cb and Cr CBs, most of the quantisation-induced compression artifacts for the higher initial QP experiments (e.g., AI QP = 37) are related to a significant distortion in terms of saturation — or colourfulness — in the chrominance data (compare the subfigures in Figure 4.12). It is important to note, however, that at the viewing distance in the subjective evaluations (i.e., $4 \times H$, where H is the height of the TV/VDU or 1.5m/59.1 inch), these chroma Cb and Cr distortions are not significantly perceptible to the human observer, especially so in the RA QP = 37 tests in which GOP-based inter coding is utilised. This fact constitutes a form of chrominance HVS-based visual masking contingent upon viewing distance, luminance/chrominance adaptation in the viewing conditions and temporal masking based on bidirectional inter prediction. This observation related to HVS-based spatiotemporal chrominance visual masking is a discovery that has not been covered in the literature and, thus, potentially gives rise to opening new lines of research related to the perceptual quantisation of chroma data.

4.4 Summary

In terms of the proposed CCCPQ technique, a novel CU level colour-based perceptually adaptive quantisation scheme is proposed for HEVC to potentially replace AdaptiveQP technique. CCCPQ accounts for the spatial activity of Y, Cb and Cr sample data, in the corresponding CBs, in order to select a more appropriate $2N \times 2N$ CU level QP during the coding process; this is achieved by employing a cross-colour channel dependency for QP selection mechanism. That is, the combined Y, Cb and Cr sample variances are computed in the constituent CBs in a $2N \times 2N$ CU; the CU-level is adjusted according the corresponding variance computations.

In terms of the proposed CBPQ technique, a novel CB-level perceptual quantisation method is proposed for HEVC for the perceptual coding of 4:4:4 data. CBPQ is designed to perceptually adjust the QP at the CB level — i.e., the QPs for the luma CB and the chroma Cb and Cr CBs — according to the variances of sample data in each CB. Recall that CCCPQ perceptually adjusts the QP for an entire $2N \times 2N$, based on the combined variances in each CB; separately adjusting the QP for the Y CB, the Cb CB and the Cr CB, as is the case with the proposed CBPQ technique, constitutes a more refined approach in terms of accounting for the spatial activity in the luma and chroma CBs. Furthermore, CBPQ exploits the chroma Cb and Cr CB-level QP offset mechanism provided by JCT-VC, which increases bitrate reductions for 4:4:4 data (in comparison with CCCPQ).

CCCPQ and CBPQ were subsequently implemented into JCT-VC HEVC HM 16.7 for the purpose of undertaking experimental evaluations. The most noteworthy coding performances attained by CCCPQ and CBPQ compared with AdaptiveQP are as follows: i) CCCPQ: a 22.5% bitrate reduction when applied to the 4:4:4 Kimono sequence using the AI encoding configuration; ii) CBPQ: a 25.4% bitrate reduction when applied to the 4:4:4 Kimono sequence using the AI encoding configuration (no differences in perceptual quality). In addition to this, the subjective evaluations reveal that visually lossless coding is achieved by the proposed techniques (compared with the raw data and also AdaptiveQP) in all of the initial AI and RA QP = 22 tests. No significant differences are recorded in terms of runtimes, which provides evidence that CCCPQ and CBPQ do not incur increased computational complexity.

Chapter 5. JND-Based Perceptually Adaptive Quantisation

Expanding on the CBPQ technique proposed in Chapter 4, in this chapter a CB-level JND-based luma and chroma perceptual quantisation technique (named Pixel-PAQ) is proposed. Pixel-PAQ [24, 25] is designed to perceptually increase the Y QP, the Cb QP and the Cr QP at the CB level in HEVC; this approach facilitates the JND-based perceptual coding of both luma and chroma data. A significant feature of Pixel-PAQ is that it extends Naccari's and Mrak's JND-based IDSQ technique in [42]; that is, the JND for chrominance data is accounted for in Pixel-PAQ (as opposed to luminance data only, which is the case with IDSQ). Accordingly, Pixel-PAQ exploits both luminance masking and chrominance masking based on spatial CSF-related luminance adaptation and chrominance adaptation.

In relation to the perceptual coding of chroma Cb and Cr data, Pixel-PAQ has the potential to considerably outperform Naccari's and Mrak's JND-based IDSQ technique in terms of bitrate reductions. According to the aforementioned chrominance CSF-related functions, Pixel-PAQ is designed to apply coarser levels of quantisation to Cb and Cr data when coding YCbCr 4:4:4 data and also 4:2:0, 4:2:2 chroma subsampled data. The proposed method is particularly effective when applied to high bit depth YCbCr 4:4:4 video data, primarily because the Cb and Cr channels in high bit depth 4:4:4 data typically contain a considerable amount of perceptual redundancy due to the higher variances in the chroma channels. Moreover, compression artifacts in high variance chroma data are not conspicuous to the HVS. Therefore, the Cb and Cr data in high bit depth YCbCr 4:4:4 video sequences can be compressed much more aggressively than the Y data.

The rest of the chapter is organised as follows. Section 5.1 includes background information relevant to Pixel-PAQ. In section 5.2, the proposed technique is presented. In section 5.3, thorough subjective and objective experimental evaluations are presented and discussed. Finally, section 5.4 concludes this chapter.

5.1 Related Background

JND-based visually lossless coding is a sophisticated approach for ascertaining the point at which compression-induced distortions become perceptually discernible to the human observer. In video and image coding applications, JND is generally defined as the maximum visibility threshold before lossy compression distortions are perceptually discernible to the HVS [1, 39]; JND has its roots in the Weber–Fechner law [89]. In the Weber–Fechner law, the JND absolute threshold (the maximum visibility threshold for visual-orientated applications) is defined as the minimum change in a physical stimulus which is perceptible if the corresponding threshold is exceeded. In relation to video and image coding applications, Fencher's law decrees that there is a mathematical relationship between the subjective sensation of a physical stimulus and the intensity of the actual physical stimulus. This implies that there exists a mathematical relationship between perceived brightness and the intensity of physical luminance. Likewise, it also implies that there is a mathematical relationship between perceived colour and the energy of photons. As such, JND-based visually lossless coding research is presently of considerable interest in the areas of video coding and image coding. Focusing on video compression in the HEVC standard, JND-based visually lossless coding techniques can profoundly reduce the perceptual redundancies that are present in raw YCbCr video data. Therefore, the number of bits required to store each pixel can be considerably reduced. As such, burdens related to non-volatile data storage, transmission and bandwidth can be reduced to an extremely high degree.

The key difference between visually lossless coding, JND-based perceptual video coding techniques and video coding based on rate-distortion theory is as follows: visually lossless coding and JND-based techniques prioritise the HVS with respect to assessing the reconstruction quality of a perceptually coded video. Note that with JND and visually lossless coding, PSNR measurements are not considered to be important in terms of quantifying the perceptual quality of a reconstructed sequence. In such cases, the PSNR metric is utilised for quantifying the degree to which PSNR values can be decreased before the associated compression-induced distortions in the coded video are perceptually discernible.

5.2 Pixel-PAQ: JND-Based Luma and Chroma Perceptual Quantisation

Pixel-PAQ extends Naccari's and Mrak's spatial CSF-related and luminance adaptation-based IDSQ JND technique in [42]. Pixel-PAQ focuses on extending IDSQ to incorporate chrominance JND in addition to accounting for high bit depth luma data and high bit depth chroma data. Both luminance masking and chrominance masking piecewise functions are employed to perceptually increase quantisation levels by virtue of JND-based modifications to the luma QStep and the chroma QSteps at the CB level. A primary objective of Pixel-PAQ is to decrease the number of perceptually insignificant non-zero luma and chroma transform coefficients. After entropy coding, the resulting coded bitstream will thus contain significantly fewer bits, thus reducing bitrate and non-volatile data storage requirements. The coarser quantisation noise induced by Pixel-PAQ is indiscernible to the human observer assuming that the luma and chroma JND visibility thresholds are not exceeded.

Naccari's and Mrak's JND-based IDSQ method is founded upon the technique in [42] by X. H. Zhang et al. In [43], the authors conclude that there is an intrinsic relationship between luminance adaptation, background luminance and the corresponding luminance data in an image. Concerning luminance adaptation, X. H. Zhang et al. assert that the contrast threshold for luminance exhibits a parabolic curve corresponding to CSF-related grey level luminance, from which a parabolic piecewise function is derived. Naccari and Mrak employ this piecewise function in IDSQ [42].

5.2.1 JND-Based Luminance Perceptual Quantisation

In Pixel-PAQ, the aforementioned parabolic piecewise function, denoted as $L(\mu_Y)$, which also constitutes the luma JND visibility threshold, is utilised as a weight to increase the luma QStep in HEVC. Function $L(\mu_Y)$ is computed in (5.1):

$$L(\mu_Y) = \begin{cases} a \cdot \left(1 - \dfrac{2\mu_Y}{2^b}\right)^d + 1, & \text{if } \mu_Y \le \dfrac{2^b}{2} \\ c \cdot \left(\dfrac{2\mu_Y}{2^b} - 1\right)^f + 1, & \text{otherwise} \end{cases} \tag{5.1}$$

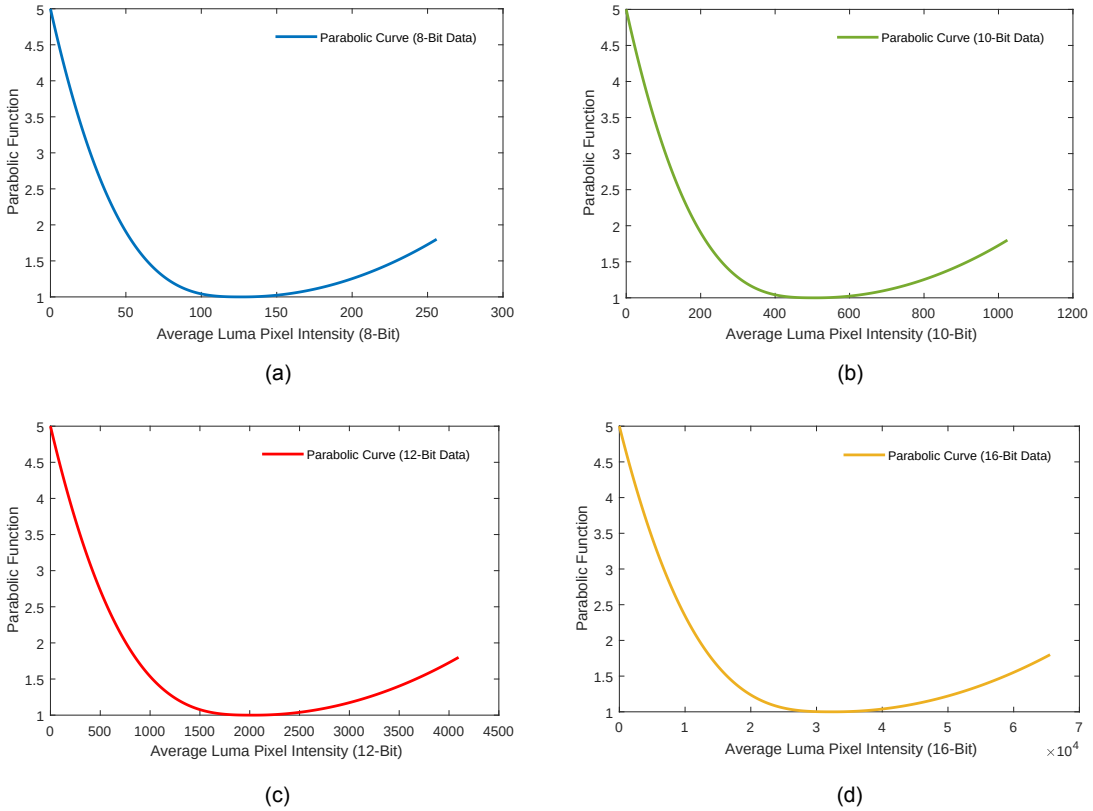


Figure 5.1: The curves derived from the parabolic function $L(\mu_Y)$ in (1). Note that the subfigures are as follows: (a) corresponds to the parabolic curve when b = 8 (8-bit luma data), (b) b = 10 (10-bit luma data), (c) b = 12 (12-bit luma data) and b = 16 (16-bit luma data). Note how, regardless of the bit depth of the luma data, the integrity of the parabolic curve is preserved.

where parameters a, c, d and f are set to values 2, 0.8, 3 and 2, respectively. These parameter values are selected by X. H. Zhang et al. in [43] to determine the shape of the spatial CSF-related luminance adaptation parabolic curve (see Figure 5.1). Recall that Naccari and Mrak adopt a very similar function to (5.1) in their IDSQ technique [42].

X. H. Zhang et al. in [43] approximate the shape of the parabola, as shown in Figure 5.1 (a), based on the luminance CSF-related psychophysical experiments conducted by Ahumada and Peterson in [70]. In contrast to the original equation for (1) in [42], the value 256 is replaced by 2^b, where superscript b denotes the bit depth of the raw sample data in a luma CB; this is to ensure that the proposed method is compatible with luma data of any bit depth. As shown in Figure 5.1, the integrity of the parabolic curve is preserved regardless of the value of b in (5.1).

Assuming that value 256 in (1) in [42] is replaced with 2^b in (5.1), $L(\mu_Y)$ can subsequently be utilised in perceptual quantisation techniques for luma data of any bit depth. Furthermore, it is important to note that the mean values for the full range of luma data for any bit depth — i.e., (0+256/2) for 8-bit data, (0+1024/2) for 10-bit data, (0+4096/2) for 12-bit data and (0+65536/2) for 16-bit data — equates to a perceptually identical shade of greyscale colour in the luma component.

In (5.1), variable μ_Y denotes the mean raw sample value in a luma CB; μ_Y is computed in (5.2):

$$\mu_Y = \frac{1}{2N \times 2N} \sum_{n=1}^{2N \times 2N} w_{Y_n} \tag{5.2}$$

where $2N \times 2N$ denotes the number of sample values in a luma CB and where variable w_Y refers to the n^{th} sample value in a luma CB. To reiterate, we compute μ_Y from the original, raw sample values at the luma CB level.

There is a binary logarithmic relationship between the QP and the QStep in URQ in HEVC; this is the case for both frame-level and CB-level luma and chroma quantisation. In the luminance JND aspect of Pixel-PAQ, the primary objective is to perceptually increase the luma QStep by weighing it with $L(\mu_Y)$. In URQ in HEVC, the luma QP (denoted as QP_Y) and the luma QStep (denoted as $QStep_Y$) are computed in (5.3) and (5.4), respectively.

$$QP_Y\left(QStep_Y\right) = \left[6 \times \log_2\left(QStep_Y\right)\right] + 4 \tag{5.3}$$

$$QStep_Y\left(QP_Y\right) = 2^{\frac{QP_Y - 4}{6}} \tag{5.4}$$

The quantisation-induced error, denoted as q_Y, after the inverse quantisation of the luma coefficients is only perceptually discernible in the reconstructed signal if it exceeds the luma JND visibility threshold $L(\mu_Y)$. Visually lossless coding is therefore achieved if $|q_Y| \leq L(\mu_Y)$.

To reiterate, the luma QStep that incurs the maximum amount of perceptually indiscernible quantisation-induced distortion is achieved by adaptively weighing $QStep_Y$ with $L(\mu_Y)$. Therefore, the CB-level JND-based perceptual luma QStep, denoted as $PStep_Y$, is quantified in (5.5).

$$PStep_Y = QStep_Y \cdot \left[L\left(\mu_Y\right) \right] \tag{5.5}$$

Accordingly, the CB-level JND-based perceptual luma QP, denoted as PQP_Y, is computed in (5.6).

$$PQP_Y\left(PStep_Y\right) = \left[6 \times \log_2\left(PStep_Y\right) \right] + 4 \tag{5.6}$$

5.2.2 JND-Based Chrominance Perceptual Quantisation

As previously mentioned in the literature review chapter, Naccari and Pereira in [67] propose a JND-based QM method for the AVC standard. The authors assume that perceptual masking is similar for both luma and chroma data based on the premise that the luminance and chrominance spatial CSFs exhibit similar properties. Therefore, Naccari and Pereira in [67] apply the same JND visibility threshold for both luma and chroma perceptual quantisation. Although the luminance spatial CSF and the chrominance spatial CSF share somewhat similar properties [90], there are obvious differences between the two, especially in relation to the comparative sensitivity of the HVS to achromatic data and chromatic data in compressed video data [91, 44].

In the proposed Pixel-PAQ method, a relatively similar piecewise function to (5.1) is utilised for the CB-level JND-based perceptual quantisation of chroma Cb and Cr data. The corresponding chrominance piecewise functions, denoted as $C_{Cb}(\mu_{Cb})$ and $C_{Cr}(\mu_{Cr})$, are based on chrominance adaptation related to chrominance spatial CSF. The mean values of $C_{Cb}(\mu_{Cb})$ and $C_{Cr}(\mu_{Cr})$ are denoted as μC_{Cb} and μC_{Cr}, respectively, which are utilised to perceptually weigh the chroma QSteps. Functions $C_{Cb}(\mu_{Cb})$ and $C_{Cr}(\mu_{Cr})$, which constitute the chroma Cb and Cr JND visibility thresholds, are computed in (5.7) and (5.8), respectively:

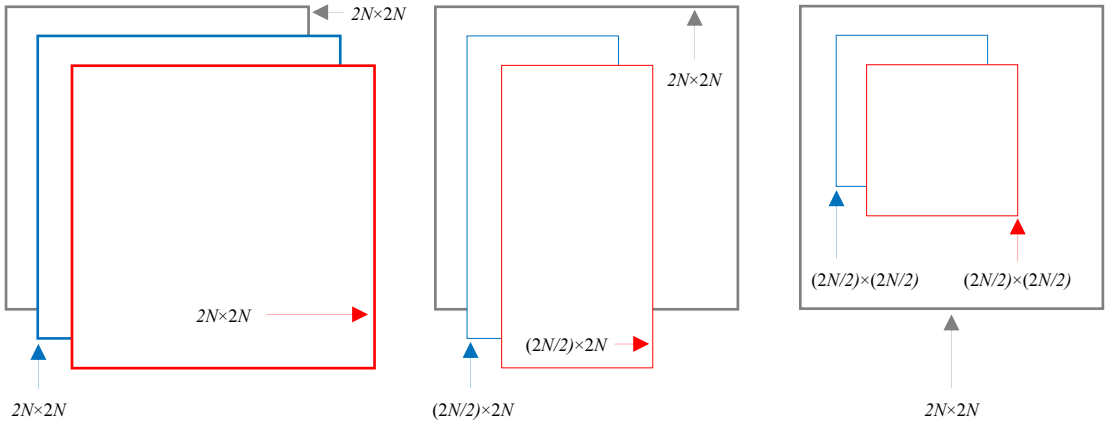
$$C_{Cb}\left(\mu_{Cb}\right)=\begin{cases}\dfrac{-\mu_{Cb}\cdot\left(g-1\right)}{h+g} & \text{if } \mu_{Cb}\le h\\ 1 & \text{if } h<\mu_{Cb}<j\\ \dfrac{\left(\mu_{Cb}-j\right)\cdot\left(k-1\right)}{\left(2^{b}-1-j\right)+1}, & \text{otherwise}\end{cases} \tag{5.7}$$

$$C_{Cr}\left(\mu_{Cr}\right)=\begin{cases}\dfrac{-\mu_{Cr}\cdot\left(g-1\right)}{h+g} & \text{if } \mu_{Cr}\le h\\ 1 & \text{if } h<\mu_{Cr}<j\\ \dfrac{\left(\mu_{Cr}-j\right)\cdot\left(k-1\right)}{\left(2^{b}-1-j\right)+1}, & \text{otherwise}\end{cases} \tag{5.8}$$

where parameters g, h, j and k are set to values 3, 85, 90 and 3, respectively. Similar to the way in which Naccari and Mrak adopt parameter values a, c, d and f in (5.1) for IDSQ (i.e., based on the work conducted by X. H. Zhang et al. in [43]), the values for parameters g, h, j and k in Pixel-PAQ are selected based on the chrominance adaptation psychophysical experiments conducted by G. Wang et al. in [44]. In [44], the authors conduct chrominance spatial CSF-based psychophysical experiments, from which the values for g, h, j and k are derived; these parameter values are adopted for utilisation in Pixel-PAQ. Note that the data in the chroma Cb and Cr channels exhibit very similar spatial properties [23, 44]. Therefore, parameter values g, h, j and k are employed in both (5.7) and (5.8), respectively. Variables μ_{Cb} and μ_{Cr} denote the mean raw chroma sample values in chroma Cb and Cr CBs, respectively; they are computed in (5.9) and (5.10), respectively:

$$\mu_{Cb}=\frac{1}{M}\sum_{m=1}^{M}h_{Cb_m} \tag{5.9}$$

$$\mu_{Cr}=\frac{1}{M}\sum_{m=1}^{M}s_{Cr_m} \tag{5.10}$$



(a) 4:4:4 YCbCr CB Sizes **(b)** 4:2:2 YCbCr CB Sizes **(c)** 4:2:0 YCbCr CB Sizes

Figure 5.2: The sizes of Y, Cb and Cr CBs in a $2N\times2N$ CU in HEVC: Y (grey), Cb (blue), Cr (red). Each subfigure specifies the size of Cb and Cr CBs for different raw video data: (a) for 4:4:4 YCbCr video data, the CB sizes for Y, Cb and Cr are all $2N\times2N$, (b) for YCbCr 4:2:2 video data, the CB sizes are as follows: $Y_{CB} = 2N\times2N$, $Cb_{CB} = (2N/2)\times2N$ and $Cr_{CB} = (2N/2)\times2N$. (c) for YCbCr 4:2:0 data, the CB sizes are as follows: $Y_{CB} = 2N\times2N$, $Cb_{CB} = (2N/2)\times(2N/2)$ and $Cr_{CB} = (2N/2)\times(2N/2)$.

where M denotes the number of sample values in the chroma Cb and Cr CBs, variable h_{Cb} refers to the m^{th} sample value in a Cb CB and variable s_{Cr} refers to the m^{th} sample value in a Cr CB. Unlike the number of sample values in Y CBs, and also due to potential chroma subsampling, M is not a fixed value. Moreover, note that Cb and Cr CBs are always identical in size regardless of the chroma sampling ratio (e.g., 4:4:4, 4:2:2 or 4:2:0) — see Figure 5.2. As is the case with QP_Y and $QStep_Y$ in (5.3) and (5.4), respectively, in URQ there is a binary logarithmic relationship between the chroma Cb and Cr QPs (denoted as QP_{Cb} and QP_{Cr}, respectively) and the chroma Cb and Cr QSteps (denoted as $QStep_{Cb}$ and $QStep_{Cr}$, respectively). Accordingly, QP_{Cb}, $QStep_{Cb}$, QP_{Cr} and $QStep_{Cr}$ are computed in (5.11)-(5.14), respectively:

$$QP_{Cb}\left(QStep_{Cb}\right) = \left\lceil 6\times\log_2\left(QStep_{Cb}\right)\right\rceil + 4 \tag{5.11}$$

$$QStep_{Cb}\left(QP_{Cb}\right) = 2^{\frac{QP_{Cb}-4}{6}} \tag{5.12}$$

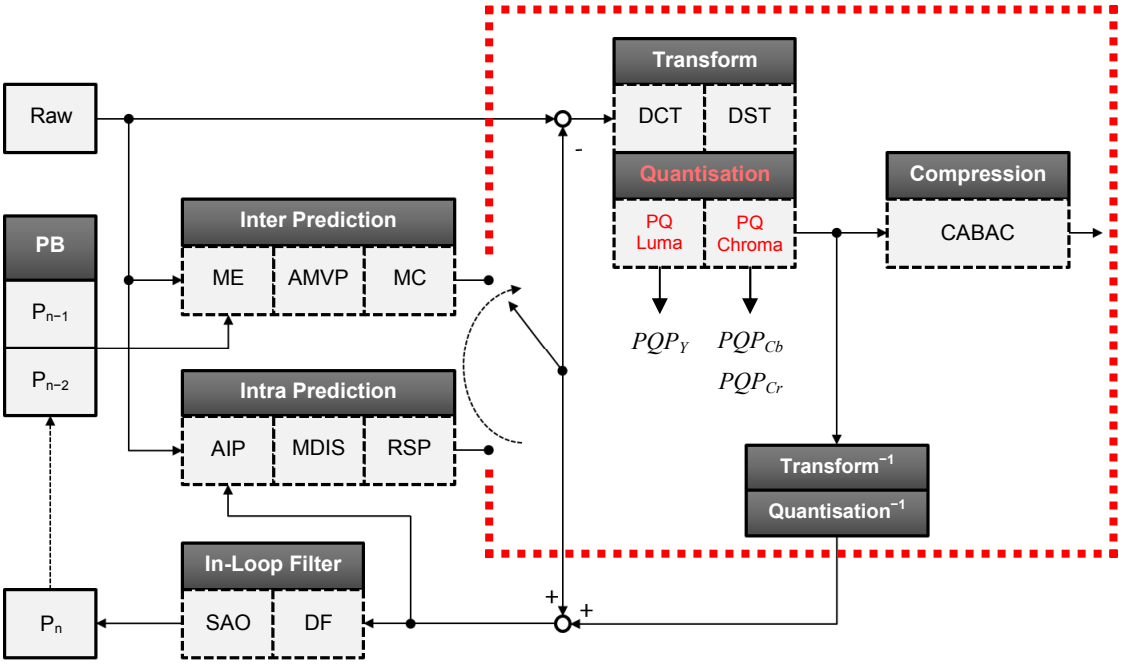


Figure 5.3: A block diagram which shows the proposed Pixel-PAQ method implemented into the HEVC HM encoder. The red dotted line and the red text indicate the areas within the HEVC coding pipeline in which the proposed method is implemented. Note that variables PQP_Y, PQP_{Cb} and PQP_{Cr} denote the perceptually adaptive QPs.

$$QP_{Cr}\left(QStep_{Cr}\right)=\left[6\times\log_2\left(QStep_{Cr}\right)\right]+4 \tag{5.13}$$

$$QStep_{Cr}\left(QP_{Cr}\right)=2^{\frac{QP_{Cr}-4}{6}} \tag{5.14}$$

Recall that the HVS is significantly more sensitive to spatial contrast in luminance data compared with the corresponding spatial contrast sensitivity response to chromatic data. This correlates with the well established fact that the HVS is considerably less sensitive to gradations — including quantisation-induced compression artifacts — in compressed chroma data. This is the main reason why chrominance data can be quantised much more aggressively, especially high variance chroma data. To reiterate, quantisation-induced compression artifacts are vastly more perceptible in reconstructed luma data; this is primarily due to the fact that the luma channel contains all of the fine details in YCbCr pictures [92, 1].

The quantisation-induced distortions, denoted as q_{Cb} and q_{Cr}, respectively, after the inverse quantisation of chroma Cb and Cr coefficients, are noticeable if they exceed the chroma Cb and Cr JND visibility thresholds $C_{Cb}(\mu_{Cb})$ and $C_{Cr}(\mu_{Cr})$, respectively. Visually lossless coding is achieved if $|q_{Cb}| \leq C_{Cb}(\mu_{Cb})$ and $|q_{Cr}| \leq C_{Cb}(\mu_{Cr})$. To attain the JND-based perceptual quantisation of chroma Cb and Cr data, $QStep_{Cb}$ and $QStep_{Cr}$ are weighed accordingly. The chroma perceptual QSteps and QPs, denoted as $PStep_{Cb}$, $PStep_{Cr}$, PQP_{Cb} and PQP_{Cr}, are computed in (5.15)-(5.18), respectively.

$$PStep_{Cb} = QStep_{Cb} \cdot \left[\mu C_{Cb}\right] \tag{5.15}$$

$$PQP_{Cb}\left(PStep_{Cb}\right) = \left[6 \times \log_2\left(PStep_{Cb}\right)\right] + 4 \tag{5.16}$$

$$PStep_{Cr} = QStep_{Cr} \cdot \left[\mu C_{Cr}\right] \tag{5.17}$$

$$PQP_{Cr}\left(PStep_{Cr}\right) = \left[6 \times \log_2\left(PStep_{Cr}\right)\right] + 4 \tag{5.18}$$

In relation to the initial QPs utilised to evaluate Pixel-PAQ (QPs 17, 22, 27, 32 and 37), the proposed method is implemented into HEVC HM by exploiting the CB-level chroma Cb and Cr QP offset signalling mechanism provided by JCT-VC [18, 19]. Therefore, the Cb and Cr QPs are perceptually increased at the CB level by offsetting them against PQP_Y. These QP and QStep offsets, denoted as OQP_{Cb}, $OStep_{Cb}$, OQP_{Cr} and $OStep_{Cr}$, respectively, are quantified in (5.19)-(5.22).

$$OQP_{Cb}\left(\mu C_{Cb}\right) = PQP_Y + \left[3\mu C_{Cb}\right] \tag{5.19}$$

$$OStep_{Cb}\left(\mu C_{Cb}\right) = 2^{\frac{PQP_Y + \left[3\mu C_{Cb}\right] - 4}{6}} \tag{5.20}$$

$$OQP_{Cr}\left(\mu C_{Cr}\right) = PQP_Y + \left[3\mu C_{Cr}\right] \tag{5.21}$$

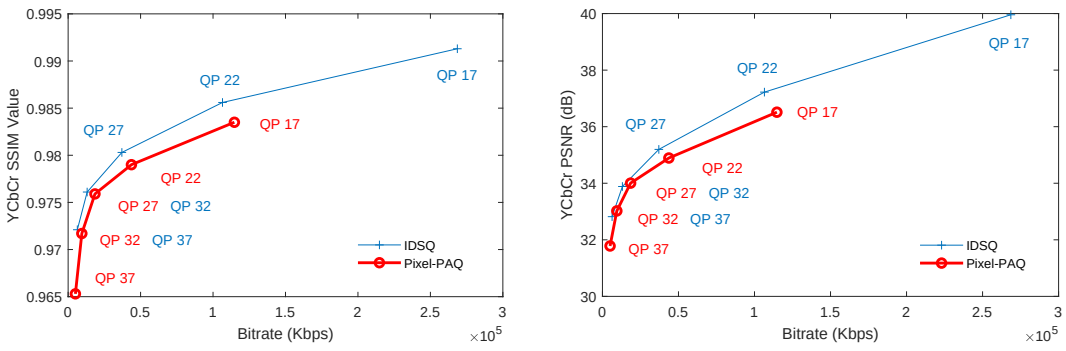
$$OStep_{Cr}\left(\mu C_{Cr}\right) = 2^{\frac{PQP_Y + \left[3\mu C_{Cr}\right] - 4}{6}} \tag{5.22}$$

Regarding the CB-level chroma Cb and Cr QP offset signalling technique present in the latest versions of JCT-VC HEVC HM [29, 30], this method is also exploited in our previously published perceptual quantisation contribution [23]; see CBPQ (Chapter 4 in this thesis). That is, we exploit the flexibility provided by JCT-VC in terms of signalling to the decoder — in the Picture Parameter Set (PPS) [93, 86] — chroma QP offsets at the CB level. The signalling of CB-level Cb and Cr QP offsets in the PPS proved to be particularly advantageous for CBPQ (Chapter 4), primarily because it allows for a straightforward encoder side implementation (see Figure 5.3). In essence, by employing this chroma QP offset scheme, all of the CB-level quantisation-related data can be efficiently transmitted to the decoder; this ensures that the perceptually compressed video is correctly decoded and reconstructed.

5.3 Experimental Evaluations, Results and Discussion

The same experimental evaluation procedure is employed in all three contribution chapters of this PhD thesis. Please refer to section 1.2 in Chapter 1 for a detailed overview of the conditions employed in the objective and subjective evaluations.

Pixel-PAQ is evaluated and compared with the IDSQ technique proposed by Naccari and Mrak in [42], which has been previously proposed for the HEVC standard. It is important to affirm that IDSQ outperforms AdaptiveQP, RDOQ and URQ in terms of bitrate reductions attained without affecting perceptual quality. Therefore, comparing Pixel-PAQ with IDSQ also implies comparing it with AdaptiveQP, RDOQ and URQ.

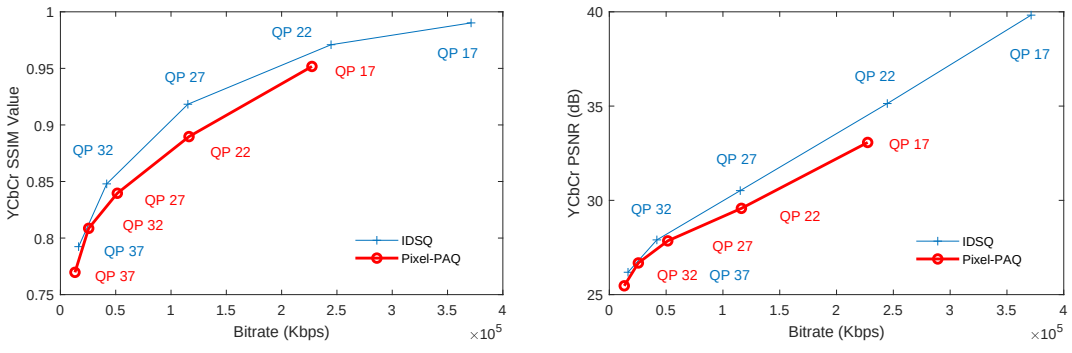


(a) (b)

Figure 5.4: Two plots which highlight the bitrate reductions attained by Pixel-PAQ compared with IDSQ. The subfigures show the bitrate reductions achieved by IDSQ on the following sequences. Subfigure (a): BirdsInCage 4:4:4 (AI - SSIM). Subfigure (b): BirdsInCage 4:4:4 (AI - PSNR).

Table 5.1: The overall bitrate reductions attained, per sequence, for the proposed Pixel-PAQ technique compared with IDSQ. The bitrates, in Kbps, are averaged over five QP data points (initial QPs 17, 22, 27, 37 and 37). The AI results are shown on the left; the RA results are shown on the right. The green text indicates superior results (i.e., lower bitrates), which are all accomplished by the proposed method.

Mean Bitrate (Kbps) – AI (YCbCr 4:2:0)			**Mean Bitrate (Kbps) – RA (YCbCr 4:2:0)**		
Sequence	Pixel-PAQ	IDSQ	**Sequence**	Pixel-PAQ	IDSQ
BirdsInCage	**32,033**	37,187	BirdsInCage	**4,934**	5,555
DuckAndLegs	**71,610**	85,760	DuckAndLegs	**17,239**	20,128
Kimono	**15,348**	18,699	Kimono	**4,519**	5,192
OldTownCross	**87,742**	99,078	OldTownCross	**24,876**	26,573
ParkScene	**20,058**	17,981	ParkScene	**2,988**	3,229
Traffic	**42,440**	47,980	Traffic	**5,845**	6,434
Mean Bitrate (Kbps) – AI (YCbCr 4:2:2)			**Mean Bitrate (Kbps) – RA (YCbCr 4:2:2)**		
Sequence	Pixel-PAQ	IDSQ	**Sequence**	Pixel-PAQ	IDSQ
BirdsInCage	**31,642**	45,420	BirdsInCage	**4,688**	7,615
DuckAndLegs	**78,478**	113,450	DuckAndLegs	**17,918**	29,195
Kimono	**15,377**	24,948	Kimono	**4,424**	6,711
OldTownCross	**89,234**	121,893	OldTownCross	**24,585**	32,838
ParkScene	**38,081**	54,772	ParkScene	**6,745**	10,147
Traffic	**43,217**	52,871	Traffic	**5,879**	7,338
Mean Bitrate (Kbps) – AI (YCbCr 4:4:4)			**Mean Bitrate (Kbps) – RA (YCbCr 4:4:4)**		
Sequence	Pixel-PAQ	IDSQ	**Sequence**	Pixel-PAQ	IDSQ
BirdsInCage	**38,279**	86,335	BirdsInCage	**5,730**	20,985
DuckAndLegs	**86,760**	157,897	DuckAndLegs	**17,917**	29,195
Kimono	**17,792**	40,907	Kimono	**20,679**	57,037
OldTownCross	**105,442**	202,409	OldTownCross	**27,040**	68,202
ParkScene	**43,017**	76,678	ParkScene	**7,409**	17,148
Traffic	**46,172**	63,806	Traffic	**6,269**	9,888



(a) (b)

Figure 5.5: Two plots which highlight the bitrate reductions attained by Pixel-PAQ compared with IDSQ. The subfigures show the bitrate reductions achieved by IDSQ on the following sequences. Subfigure (a): DuckAndLegs 4:4:4 (AI - SSIM). Subfigure (b): DuckAndLegs 4:4:4 (AI - PSNR).

Table 5.2: The bitrate reduction percentages (in green text), per sequence, attained for the proposed Pixel-PAQ technique compared with IDSQ. In addition, the decreased reconstruction quality (per channel) for sequences coded by Pixel-PAQ, as quantified by SSIM percentage decreases, are tabulated. The bitrate reductions are averaged over five QP data points (initial QPs 17, 22, 27, 37 and 37). The AI results are shown on the left; the RA results are shown on the right.

Mean Bitrate (%) Per Sequence and SSIM (%) Per Channel: Pixel-PAQ Versus IDSQ (YCbCr 4:2:0)

Sequence	All Intra				Random Access			
	Bitrate	**Y SSIM**	**Cb SSIM**	**Cr SSIM**	**Bitrate**	**Y SSIM**	**Cb SSIM**	**Cr SSIM**
BirdsInCage	**−13.9**	0.0	−0.6	−0.4	**−11.2**	0.0	−0.3	−0.3
DuckAndLegs	**−16.5**	0.0	−5.3	−4.6	**−14.4**	0.0	−3.9	−4.1
Kimono	**−17.9**	0.0	−1.0	−0.7	**−13.0**	0.0	−0.5	−0.4
OldTownCross	**−11.4**	0.0	−1.9	−1.4	**−6.4**	0.0	−0.8	−1.0
ParkScene	**−10.4**	0.0	−2.2	−1.1	**−7.5**	0.0	−1.8	−0.9
Traffic	**−11.5**	0.0	−2.0	−1.2	**−9.2**	0.0	−1.6	−1.0

Mean Bitrate (%) Per Sequence and SSIM (%) Per Channel: Pixel-PAQ Versus IDSQ (YCbCr 4:2:2)

Sequence	All Intra				Random Access			
	Bitrate	**Y SSIM**	**Cb SSIM**	**Cr SSIM**	**Bitrate**	**Y SSIM**	**Cb SSIM**	**Cr SSIM**
BirdsInCage	**−30.3**	0.0	−1.4	−0.8	**−38.4**	0.0	−0.6	−0.4
DuckAndLegs	**−30.8**	0.0	−9.6	−7.7	**−38.6**	0.0	−7.3	−6.9
Kimono	**−38.4**	0.0	−2.4	−0.9	**−34.1**	0.0	−0.6	−0.5
OldTownCross	**−26.8**	0.1	−4.5	−3.3	**−25.1**	0.0	−1.8	−1.4
ParkScene	**−30.5**	0.1	−5.4	−3.1	**−33.5**	0.0	−3.0	−2.5
Traffic	**−18.3**	0.0	−3.2	−1.7	**−19.9**	0.0	−2.6	−1.5

Mean Bitrate (%) Per Sequence and SSIM (%) Per Channel: Pixel-PAQ Versus IDSQ (YCbCr 4:4:4)

Sequence	All Intra				Random Access			
	Bitrate	**Y SSIM**	**Cb SSIM**	**Cr SSIM**	**Bitrate**	**Y SSIM**	**Cb SSIM**	**Cr SSIM**
BirdsInCage	**−55.7**	0.0	−3.3	−0.8	**−72.7**	0.0	−1.4	−0.5
DuckAndLegs	**−45.1**	0.0	−14.6	−9.7	**−63.7**	0.0	−13.2	−9.0
Kimono	**−56.5**	0.0	−4.0	−1.2	**−62.6**	0.0	−1.5	−0.5
OldTownCross	**−47.9**	0.0	−9.2	−5.3	**−60.4**	0.0	−4.5	−2.6
ParkScene	**−43.9**	0.0	−6.7	−3.9	**−56.8**	0.0	−4.0	−2.9
Traffic	**−27.6**	0.0	−3.6	−1.9	**−36.6**	0.0	−2.8	−1.6

The most outstanding result is achieved on the BirdsInCage 4:4:4 sequence in the RA QP = 22 test (see Table 5.1 and Figure 5.5). In this particular test and compared with IDSQ, a 75% bitrate reduction is achieved by Pixel-PAQ; this averages out at a 72.7% bitrate reduction over five initial QP values (i.e., QPs 17, 22, 27, 32 and 37). In the RA QP = 22 test, the following bitrate reduction is attained: 4,395 Kbps (Pixel-PAQ) versus 17,760 Kbps (IDSQ) for 600 frames. Furthermore, as confirmed in the subjective evaluations, visually lossless coding is achieved (see Figure 5.6).

(a) Pixel-PAQ Coded Inter-Frame (RA, QP = 22): YCbCr PSNR = 34.5736 and YCbCr SSIM = 0.9777

(b) Raw Data

Figure 5.6: A frame from the BirdsInCage 4:4:4 sequence. Subfigure (a) is a Pixel-PAQ coded inter-frame from this sequence (RA QP = 22 test). Subfigure (b) is the corresponding raw data. In spite of the extremely high bitrate reduction of 75% for this particular test (see Table 5.1 and Table 5.2), the Pixel-PAQ coded sequence in (a) is perceptually indistinguishable from the raw data in (b); this is also confirmed in the subjective evaluations.

Table 5.3: The bitrate reduction percentages (in green text), per sequence, attained for the proposed Pixel-PAQ technique compared with IDSQ. In addition, the decreased reconstruction quality (per channel) for sequences coded by Pixel-PAQ, as quantified by PSNR percentage decreases, are tabulated. The bitrate reductions are averaged over five QP data points (initial QPs 17, 22, 27, 37 and 37). The AI results are shown on the left; the RA results are shown on the right.

Mean Bitrate (%) Per Sequence and PSNR (%) Per Channel: Pixel-PAQ Versus IDSQ (YCbCr 4:2:0)

Sequence	All Intra				Random Access			
	Bitrate	**Y PSNR**	**Cb PSNR**	**Cr PSNR**	**Bitrate**	**Y PSNR**	**Cb PSNR**	**Cr PSNR**
BirdsInCage	**−13.9**	0.0	−4.5	−5.3	**−11.2**	0.0	−3.2	−4.7
DuckAndLegs	**−16.5**	0.0	−7.5	−7.3	**−14.4**	0.0	−4.6	−5.3
Kimono	**−17.9**	0.0	−5.0	−5.0	**−13.0**	0.0	−3.5	−3.6
OldTownCross	**−11.4**	0.0	−4.8	−5.2	**−6.4**	0.0	−2.9	−4.0
ParkScene	**−10.4**	0.0	−6.9	−5.4	**−7.5**	0.0	−5.6	−4.4
Traffic	**−11.5**	0.0	−6.6	−6.1	**−9.2**	−0.1	−5.1	−5.2

Mean Bitrate (%) Per Sequence and PSNR (%) Per Channel: Pixel-PAQ Versus IDSQ (YCbCr 4:2:2)

Sequence	All Intra				Random Access			
	Bitrate	**Y PSNR**	**Cb PSNR**	**Cr PSNR**	**Bitrate**	**Y PSNR**	**Cb PSNR**	**Cr PSNR**
BirdsInCage	**−30.3**	0.0	−5.7	−6.7	**−38.4**	0.0	−3.6	−5.8
DuckAndLegs	**−30.8**	0.0	−9.3	−8.9	**−38.6**	0.1	−5.6	−6.3
Kimono	**−38.4**	0.0	−6.1	−5.7	**−34.1**	0.0	−3.6	−4.3
OldTownCross	**−26.8**	0.1	−6.6	−6.5	**−25.1**	0.0	−3.5	−4.3
ParkScene	**−30.5**	0.0	−8.0	−7.0	**−33.5**	0.0	−5.1	−5.3
Traffic	**−18.3**	0.0	−9.3	−8.5	**−19.9**	0.0	−7.3	−7.1

Mean Bitrate (%) Per Sequence and PSNR (%) Per Channel: Pixel-PAQ Versus IDSQ (YCbCr 4:4:4)

Sequence	All Intra				Random Access			
	Bitrate	**Y PSNR**	**Cb PSNR**	**Cr PSNR**	**Bitrate**	**Y PSNR**	**Cb PSNR**	**Cr PSNR**
BirdsInCage	**−55.7**	0.0	−5.6	−4.7	**−72.7**	0.1	−3.2	−3.5
DuckAndLegs	**−45.1**	−0.1	−10.7	−9.8	**−63.7**	0.0	−6.3	−6.3
Kimono	**−56.5**	0.1	−6.0	−4.7	**−62.6**	0.0	−3.1	−3.2
OldTownCross	**−47.9**	0.1	−7.6	−6.5	**−60.4**	0.1	−4.2	−3.7
ParkScene	**−43.9**	0.1	−7.8	−6.8	**−56.8**	0.1	−4.6	−4.9
Traffic	**−27.6**	0.0	−8.7	−8.2	**−36.6**	0.1	−6.3	−6.7

Compared with IDSQ, the reconstruction quality of the chroma data in the Pixel-PAQ coded sequences is significantly inferior due to the JND-based chrominance masking. However, according to the subjective evaluation results, these decreases in chroma reconstruction quality proved to be imperceptible to the HVS in many cases (see Figure 5.6). The reconstruction of luma data is not affected because the JND visibility threshold has already been reached as a result of the computations in (5.1). The mean SSIM results (per sequence, per QP) are recorded in Table 5.4 and Table 5.5.

Table 5.4: The 'per sequence' SSIM results (AI) for 'Pixel-PAQ versus the raw data' compared with 'IDSQ versus the raw data' (initial QPs 17, 22, 27, 32 and 37). The superior SSIM results (IDSQ) are shown in green text.

Mean SSIM Values (Per Sequence, Per QP): Pixel-PAQ Versus IDSQ (YCbCr 4:2:0) – All Intra

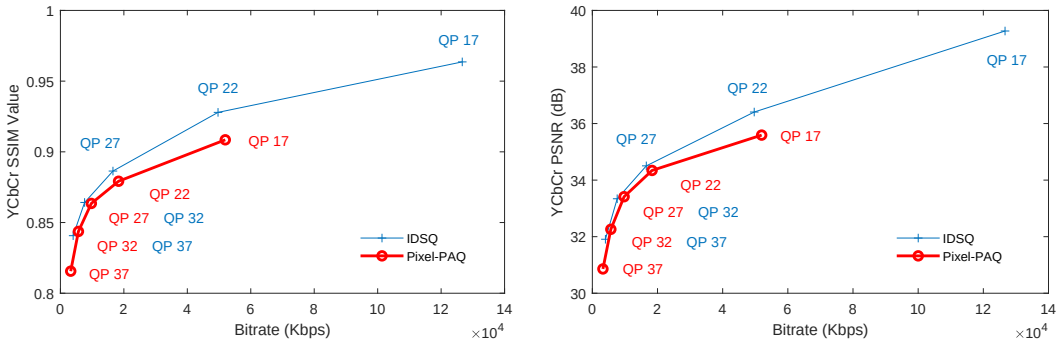
Sequence	Pixel-PAQ					IDSQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	0.9911	0.9884	0.9852	0.9821	0.9768	**0.9934**	**0.9907**	**0.9879**	**0.9850**	**0.9813**
DuckAndLegs	0.9543	0.9320	0.9005	0.8765	0.8369	**0.9840**	**0.9613**	**0.9195**	**0.8884**	**0.8509**
Kimono	0.9415	0.9227	0.9069	0.8926	0.8662	**0.9620**	**0.9396**	**0.9235**	**0.9063**	**0.8853**
OldTownCross	0.9328	0.9015	0.8534	0.8222	0.7811	**0.9625**	**0.9157**	**0.8660**	**0.8337**	**0.7958**
ParkScene	0.9585	0.9399	0.9138	0.8803	0.8292	**0.9722**	**0.9555**	**0.9302**	**0.8937**	**0.8463**
Traffic	0.9665	0.9461	0.9226	0.8953	0.8493	**0.9793**	**0.9615**	**0.9411**	**0.9128**	**0.8754**

Mean SSIM Values (Per Sequence, Per QP): Pixel-PAQ Versus IDSQ (YCbCr 4:2:2) – All Intra

Sequence	Pixel-PAQ					IDSQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	0.9890	0.9860	0.9823	0.9763	0.9641	**0.9927**	**0.9893**	**0.9860**	**0.9828**	**0.9778**
DuckAndLegs	0.9454	0.9021	0.8668	0.8357	0.7911	**0.9852**	**0.9617**	**0.9098**	**0.8610**	**0.8160**
Kimono	0.9194	0.8990	0.8814	0.8574	0.8128	**0.9608**	**0.9283**	**0.9026**	**0.8836**	**0.8570**
OldTownCross	0.9021	0.8626	0.8166	0.7762	0.7107	**0.9625**	**0.9058**	**0.8362**	**0.8007**	**0.7605**
ParkScene	0.9012	0.8682	0.8383	0.8000	0.7430	**0.9596**	**0.9146**	**0.8676**	**0.8308**	**0.7879**
Traffic	0.9686	0.9506	0.9206	0.8713	0.7864	**0.9814**	**0.9675**	**0.9456**	**0.9111**	**0.8569**

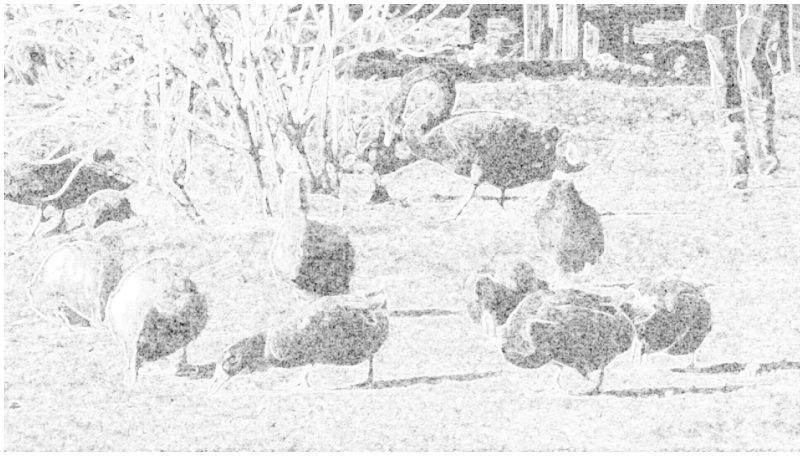
Mean SSIM Values (Per Sequence, Per QP): Pixel-PAQ Versus IDSQ (YCbCr 4:4:4) – All Intra

Sequence	Pixel-PAQ					IDSQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	0.9835	0.9790	0.9759	0.9717	0.9653	**0.9913**	**0.9856**	**0.9803**	**0.9761**	**0.9721**
DuckAndLegs	0.9517	0.8897	0.8396	0.8086	0.7697	**0.9902**	**0.9709**	**0.9183**	**0.8479**	**0.7924**
Kimono	0.9085	0.8792	0.8635	0.8437	0.8156	**0.9636**	**0.9279**	**0.8864**	**0.8642**	**0.8407**
OldTownCross	0.8864	0.8001	0.7515	0.7205	0.6823	**0.9686**	**0.9135**	**0.8122**	**0.7391**	**0.6992**
ParkScene	0.9001	0.8554	0.8268	0.7940	0.7527	**0.9639**	**0.9212**	**0.8623**	**0.8197**	**0.7774**
Traffic	0.9701	0.9535	0.9282	0.8888	0.8271	**0.9826**	**0.9686**	**0.9473**	**0.9157**	**0.8682**

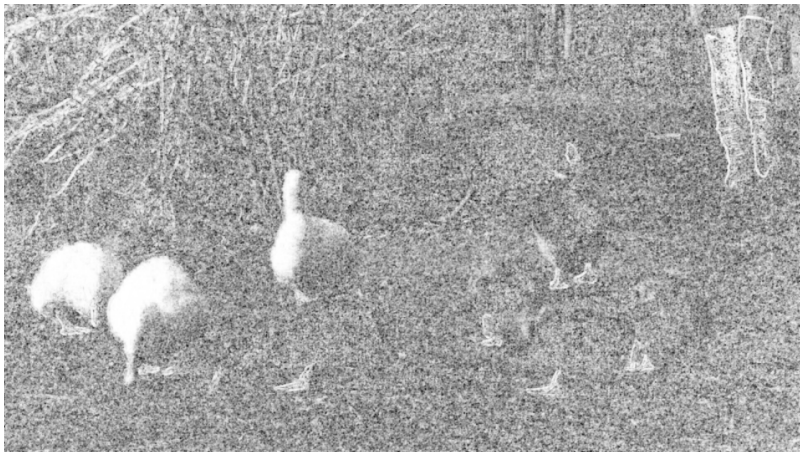


(a) (b)

Figure 5.7: Two plots which highlight the inferior mathematical reconstruction quality of Pixel-PAQ coded sequences versus IDSQ coded sequences, over five QP data points (i.e., QPs 17, 22, 27, 32 and 37), as quantified by the SSIM metric. Subfigure (a): Kimono 4:4:4 (AI - SSIM). Subfigure (b): Kimono 4:4:4 (AI - PSNR).



(a) Luma Channel



(b) Chroma Cb Channel

(c) Chroma Cr Channel

Figure 5.8: The SSIM Index Map (structural reconstruction errors) of a Pixel-PAQ coded intra-frame (AI QP = 22 test) versus the raw data (DuckAndLegs 4:4:4 sequence). In subfigures (a), (b) and (c), respectively, the luma (Y), chroma Cb and chroma Cr structural reconstruction errors are shown separately.

(a) Pixel-PAQ Coded Intra-Frame (AI, QP = 22): YCbCr PSNR = 29.5754 and YCbCr SSIM = 0.8897

(b) Raw Data

Figure 5.9: A frame from the DuckAndLegs 4:4:4 sequence. Subfigure (a) is a Pixel-PAQ coded intra-frame from this sequence (AI QP = 22 test). Subfigure (b) is the corresponding raw data. Note that, despite the poor mathematical reconstruction quality of the data in the chroma Cb and Cr channels, as quantified by SSIM (see Figure 5.8), the Pixel-PAQ coded sequence in (a) is perceptually indistinguishable from the raw data in (b); this is confirmed in the subjective evaluations.

Table 5.5: The 'per sequence' SSIM results (RA) for 'Pixel-PAQ versus the raw data' compared with 'IDSQ versus the raw data' (initial QPs 17, 22, 27, 32 and 37). The superior SSIM results are shown in green text.

Mean SSIM Values (Per Sequence, Per QP): Pixel-PAQ Versus IDSQ (YCbCr 4:2:0) – Random Access

Sequence	Pixel-PAQ					IDSQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	0.9894	0.9883	0.9860	0.9832	0.9782	**0.9905**	**0.9898**	**0.9882**	**0.9857**	**0.9825**
DuckAndLegs	0.9301	0.9060	0.8897	0.8666	0.8250	**0.9489**	**0.9196**	**0.9035**	**0.8770**	**0.8381**
Kimono	0.9224	0.9136	0.9002	0.8836	0.8584	**0.9326**	**0.9251**	**0.9120**	**0.8947**	**0.8749**
OldTownCross	0.8968	0.8565	0.8445	0.8275	0.7930	**0.9032**	**0.8640**	**0.8543**	**0.8374**	**0.8056**
ParkScene	0.9515	0.9361	0.9112	0.8780	0.8284	**0.9613**	**0.9480**	**0.9245**	**0.8909**	**0.8459**
Traffic	0.9650	0.9463	0.9267	0.9025	0.8602	**0.9735**	**0.9573**	**0.9408**	**0.9167**	**0.8838**

Mean SSIM Values (Per Sequence, Per QP): Pixel-PAQ Versus IDSQ (YCbCr 4:2:2) – Random Access

Sequence	Pixel-PAQ					IDSQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	0.9867	0.9856	0.9830	0.9780	0.9673	**0.9883**	**0.9869**	**0.9856**	**0.9831**	**0.9789**
DuckAndLegs	0.9028	0.8794	0.8591	0.8292	0.7856	**0.9504**	**0.8954**	**0.8797**	**0.8512**	**0.8065**
Kimono	0.9013	0.8954	0.8766	0.8538	0.8164	**0.9096**	**0.9016**	**0.8907**	**0.8734**	**0.8492**
OldTownCross	0.8538	0.8171	0.8034	0.7798	0.7324	**0.8796**	**0.8220**	**0.8145**	**0.7977**	**0.7646**
ParkScene	0.8827	0.8665	0.8421	0.8079	0.7535	**0.8969**	**0.8816**	**0.8625**	**0.8322**	**0.7919**
Traffic	0.9656	0.9519	0.9276	0.8844	0.8108	**0.9731**	**0.9634**	**0.9460**	**0.9171**	**0.8693**

Mean SSIM Values (Per Sequence, Per QP): Pixel-PAQ Versus IDSQ (YCbCr 4:4:4) – Random Access

Sequence	Pixel-PAQ					IDSQ				
	QP 17	**QP 22**	**QP 27**	**QP 32**	**QP 37**	**QP 17**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	0.9783	0.9777	0.9759	0.9725	0.9666	**0.9831**	**0.9783**	**0.9774**	**0.9756**	**0.9724**
DuckAndLegs	0.8766	0.8531	0.8313	0.7998	0.7586	**0.9636**	**0.9030**	**0.8514**	**0.8237**	**0.7790**
Kimono	0.8777	0.8702	0.8574	0.8386	0.8127	**0.9122**	**0.8761**	**0.8673**	**0.8522**	**0.8316**
OldTownCross	0.7847	0.7504	0.7398	0.7205	0.6868	**0.8904**	**0.7492**	**0.7467**	**0.7326**	**0.7035**
ParkScene	0.8666	0.8534	0.8303	0.7994	0.7581	**0.9102**	**0.8637**	**0.8473**	**0.8204**	**0.7825**
Traffic	0.9653	0.9533	0.9322	0.8983	0.8433	**0.9718**	**0.9620**	**0.9458**	**0.9191**	**0.8771**

As shown in Figure 5.8, the structural reconstruction errors are concentrated mostly in the high variance regions in the Y, Cb and Cr channels. The HVS is less capable of detecting quantisation-induced compression artifacts in high spatial variance regions of compressed luma and chroma data [23]. Therefore, in spite of the reconstruction errors shown in Figure 5.8, visually lossless coding is attained by Pixel-PAQ in both the AI QP = 22 test — and also RA QP = 22 test (see the SSIM plot in Figure 5.7) — on the DuckAndLegs 4:4:4 sequence. This is confirmed in the subjective evaluations; see Figure 5.9 (a) versus (b) for a comparison.

Table 5.6: The 'per sequence' PSNR (dB) results (AI) for 'Pixel-PAQ versus the raw data' compared with 'IDSQ versus the raw data' (initial QPs 17, 22, 27, 32 and 37). The superior SSIM results are shown in green text.

Mean PSNR (dB) Per Sequence, Per QP: Pixel-PAQ Versus IDSQ (YCbCr 4:2:0) – All Intra

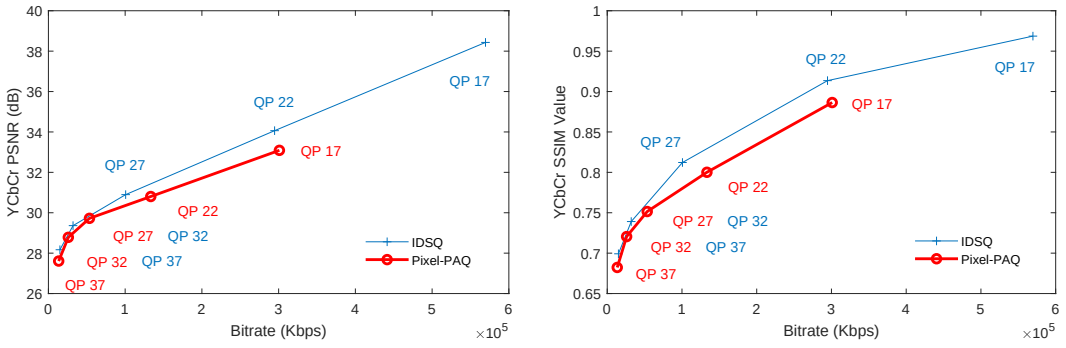
Sequence	Pixel-PAQ					IDSQ				
	QP 17	QP 22	QP 27	QP 32	QP 37	QP 17	QP 22	QP 27	QP 32	QP 37
BirdsInCage	39.1038	37.5180	36.0085	34.8799	33.3267	**40.9787**	**38.8655**	**37.1534**	**35.7128**	**34.2626**
DuckAndLegs	33.8655	31.7584	29.8392	28.6015	26.9424	**38.7371**	**34.8509**	**31.3556**	**29.3250**	**27.5127**
Kimono	37.3666	35.8659	34.5481	33.5125	31.8994	**39.6294**	**37.5313**	**35.9396**	**34.4053**	**32.7964**
OldTownCross	35.6716	33.8187	31.8953	30.6139	29.0512	**38.6377**	**35.0655**	**32.7264**	**31.1386**	**29.5463**
ParkScene	36.8713	34.6698	32.5783	30.7900	28.7431	**39.9746**	**36.9332**	**34.0121**	**31.4805**	**29.2620**
Traffic	37.2568	35.2647	33.1950	31.5932	29.5243	**40.3068**	**37.2091**	**34.8116**	**32.5567**	**30.4404**

Mean PSNR (dB) Per Sequence, Per QP: Pixel-PAQ Versus IDSQ (YCbCr 4:2:2) – All Intra

Sequence	Pixel-PAQ					IDSQ				
	QP 17	QP 22	QP 27	QP 32	QP 37	QP 17	QP 22	QP 27	QP 32	QP 37
BirdsInCage	38.0980	36.6285	35.2130	33.6314	31.6704	**40.6018**	**38.2948**	**36.5251**	**35.1635**	**33.6036**
DuckAndLegs	33.0379	30.3503	28.6050	27.2242	25.8192	**38.6569**	**34.5193**	**30.7285**	**28.4722**	**26.6346**
Kimono	36.3071	35.0233	33.7931	32.3663	30.5674	**39.3181**	**36.7747**	**35.2219**	**33.8264**	**32.1149**
OldTownCross	34.2723	32.5675	30.9991	29.6107	27.9068	**38.4001**	**34.4299**	**31.9390**	**30.5069**	**28.9463**
ParkScene	34.6031	32.5660	30.8475	29.2602	27.6884	**38.7150**	**35.2745**	**32.7290**	**30.6848**	**28.8019**
Traffic	37.5523	34.9795	32.4327	29.9572	27.5877	**40.7200**	**37.6387**	**34.6954**	**31.9075**	**29.2406**

Mean PSNR (dB) Per Sequence, Per QP: Pixel-PAQ Versus IDSQ (YCbCr 4:4:4) – All Intra

Sequence	Pixel-PAQ					IDSQ				
	QP 17	QP 22	QP 27	QP 32	QP 37	QP 17	QP 22	QP 27	QP 32	QP 37
BirdsInCage	36.5128	34.8927	34.0048	33.0221	31.7821	**39.9575**	**37.2223**	**35.1969**	**33.8851**	**32.8186**
DuckAndLegs	33.0709	29.5754	27.8507	26.6800	25.4676	**39.8238**	**35.1345**	**30.5158**	**27.9026**	**26.1863**
Kimono	35.5917	34.3420	33.4135	32.2602	30.8592	**39.2723**	**36.4072**	**34.5038**	**33.3406**	**31.9085**
OldTownCross	33.0896	30.8036	29.7243	28.7834	27.6085	**38.4316**	**34.0635**	**30.8971**	**29.3628**	**28.1692**
ParkScene	34.6143	32.5565	30.9406	29.4304	27.9728	**38.9249**	**35.4066**	**32.7230**	**30.7505**	**28.8994**
Traffic	37.8942	35.3045	32.8423	30.4738	28.2057	**41.0563**	**37.8411**	**34.8627**	**32.1028**	**29.4800**



(a) (b)

Figure 5.10: Two plots which highlight the inferior mathematical reconstruction quality of Pixel-PAQ coded sequences versus IDSQ coded sequences, over five QP data points (i.e., QPs 17, 22, 27, 32 and 37), using the SSIM metric. Subfigure (a): OldTownCross 4:4:4 (AI - SSIM). Subfigure (b): OldTownCross 4:4:4 (AI - PSNR).

Table 5.7: The 'per sequence' PSNR (dB) results (RA) for 'Pixel-PAQ versus the raw data' compared with 'IDSQ versus the raw data' (initial QPs 17, 22, 27, 32 and 37). The superior SSIM results are shown in green text.

Mean PSNR (dB) Per Sequence, Per QP: Pixel-PAQ Versus IDSQ (YCbCr 4:2:0) – Random Access

Sequence	Pixel-PAQ					IDSQ				
	QP 17	QP 22	QP 27	QP 32	QP 37	QP 17	QP 22	QP 27	QP 32	QP 37
BirdsInCage	38.0755	37.3938	36.3239	35.1878	33.6965	**38.7037**	**38.1668**	**37.2331**	**35.9925**	**34.6063**
DuckAndLegs	31.6926	30.2049	29.0677	27.8971	26.3244	**33.4621**	**31.4326**	**30.0737**	**28.4649**	**26.8022**
Kimono	36.1023	35.1974	34.0136	32.6811	31.1212	**37.2039**	**36.2523**	**34.8071**	**33.2019**	**31.6658**
OldTownCross	34.0041	32.3403	31.6902	30.8602	29.4741	**34.5554**	**32.8560**	**32.2676**	**31.3361**	**29.9690**
ParkScene	35.9882	34.2629	32.4221	30.7282	28.8326	**37.7667**	**35.7195**	**33.4312**	**31.3018**	**29.3340**
Traffic	36.6152	34.9042	33.2256	31.7335	29.8033	**38.3918**	**36.3045**	**34.3904**	**32.5393**	**30.6701**

Mean PSNR (dB) Per Sequence, Per QP: Pixel-PAQ Versus IDSQ (YCbCr 4:2:2) – Random Access

Sequence	Pixel-PAQ					IDSQ				
	QP 17	QP 22	QP 27	QP 32	QP 37	QP 17	QP 22	QP 27	QP 32	QP 37
BirdsInCage	37.0162	36.4532	35.5006	34.0949	32.1827	**37.8322**	**36.9894**	**36.4115**	**35.3608**	**33.9560**
DuckAndLegs	30.4312	29.1533	27.9895	26.7018	25.3548	**33.1084**	**30.2635**	**29.2285**	**27.7411**	**26.0493**
Kimono	35.2384	34.4255	33.2542	31.8061	30.1685	**35.9042**	**35.2015**	**34.1193**	**32.6956**	**31.1283**
OldTownCross	32.7634	31.4985	30.8881	29.9111	28.5249	**33.5117**	**31.8896**	**31.4928**	**30.6906**	**29.3754**
ParkScene	33.4226	32.1890	30.8004	29.3647	27.8868	**34.6036**	**33.4481**	**32.0739**	**30.4624**	**28.7974**
Traffic	36.6355	34.7826	32.6703	30.4097	28.1447	**38.3053**	**36.4386**	**34.3156**	**32.0356**	**29.5884**

Mean PSNR (dB) Per Sequence, Per QP: Pixel-PAQ Versus IDSQ (YCbCr 4:4:4) – Random Access

Sequence	Pixel-PAQ					IDSQ				
	QP 17	QP 22	QP 27	QP 32	QP 37	QP 17	QP 22	QP 27	QP 32	QP 37
BirdsInCage	34.7390	34.5736	34.1050	33.2660	32.1066	**36.5062**	**34.7232**	**34.4621**	**33.9179**	**33.0211**
DuckAndLegs	28.9236	28.0381	27.0870	25.9042	24.6768	**33.6996**	**29.5056**	**27.9400**	**26.7760**	**25.2985**
Kimono	34.3269	33.7630	32.8515	31.6309	30.2587	**35.6606**	**34.1753**	**33.4081**	**32.2495**	**30.9117**
OldTownCross	30.7690	30.0597	29.6798	29.0180	28.0228	**33.2456**	**30.1151**	**29.9754**	**29.4715**	**28.5145**
ParkScene	33.1118	32.1325	30.8363	29.4403	28.0782	**34.8850**	**32.9975**	**31.8978**	**30.4856**	**28.8733**
Traffic	36.5840	34.8858	32.9088	30.8086	28.6553	**38.2163**	**36.2828**	**34.2486**	**32.0847**	**29.8005**

Identical in context to the SSIM results shown in Table 5.2 and Table 5.3, and as expected, the PSNR results attained for Pixel-PAQ are inferior to those achieved by IDSQ in all tests (see Table 5.4 and Table 5.5). This equates to the fact that both PSNR and SSIM do not correspond to how the Pixel-PAQ coded sequences are visually perceived in realistic viewing situations. For example, in the RA QP = 22 test conducted on the DuckAndLegs 4:4:4 sequence, the reconstruction quality measurement achieved, as quantified by PSNR, is approximately 28.04 dB (see Table 5.6). However, as confirmed in the subjective evaluations, visually lossless coding is still achieved at this extremely low PSNR value.

Table 5.8: The MOS results, rounded to the nearest integer, of four participants in the subjective evaluations (Pixel-PAQ versus IDSQ).

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – Pixel-PAQ versus IDSQ

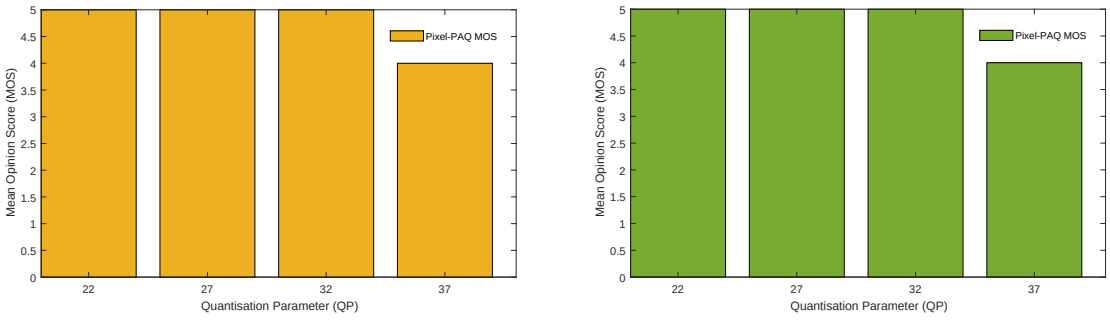
Sequence	YCbCr 4:2:0 All Intra				YCbCr 4:2:0 Random Access			
	QP 22	**QP 27**	**QP 32**	**QP 37**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	5	4	4	3	5	5	5	5
DuckAndLegs	5	5	5	5	5	5	5	5
Kimono	5	5	5	4	5	5	5	4
OldTownCross	5	5	4	4	5	5	5	4
ParkScene	5	4	3	2	5	5	5	4
Traffic	5	5	5	4	5	5	5	5

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – Pixel-PAQ versus IDSQ

Sequence	YCbCr 4:2:2 All Intra				YCbCr 4:2:2 Random Access			
	QP 22	**QP 27**	**QP 32**	**QP 37**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	5	5	4	2	5	5	4	3
DuckAndLegs	5	5	4	3	5	5	4	3
Kimono	5	5	4	3	5	5	4	2
OldTownCross	5	4	3	2	5	5	4	3
ParkScene	5	5	3	2	5	5	4	2
Traffic	5	5	3	2	5	5	5	4

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – Pixel-PAQ versus IDSQ

Sequence	YCbCr 4:4:4 All Intra				YCbCr 4:4:4 Random Access			
	QP 22	**QP 27**	**QP 32**	**QP 37**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	5	5	4	3	5	5	5	4
DuckAndLegs	5	5	5	5	5	5	5	5
Kimono	5	5	5	4	5	5	5	4
OldTownCross	5	5	4	4	5	5	4	4
ParkScene	5	5	4	3	5	5	4	4
Traffic	5	5	4	3	5	5	4	4



(a) (b)

Figure 5.11: Two Mean Opinion Score (MOS) bar graphs. Subfigure (a) shows the MOS for Pixel-PAQ versus IDSQ on the BirdsInCage 4:4:4 10-bit sequence using the RA configuration. Subfigure (b) shows the MOS for Pixel-PAQ versus the raw video data on the Traffic 4:4:4 10-bit sequence.

Table 5.9: The MOS results, rounded to the nearest integer, of four participants in the subjective evaluations for (Pixel-PAQ versus raw video data).

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – Pixel-PAQ versus Raw Data

Sequence	YCbCr 4:2:0 All Intra				YCbCr 4:2:0 Random Access			
	QP 22	**QP 27**	**QP 32**	**QP 37**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	5	3	2	2	5	4	4	3
DuckAndLegs	5	4	3	2	5	5	4	4
Kimono	5	4	3	2	5	4	3	2
OldTownCross	5	4	3	2	5	5	4	3
ParkScene	5	3	2	1	5	3	3	2
Traffic	5	5	4	2	5	5	4	4

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – Pixel-PAQ versus Raw Data

Sequence	YCbCr 4:2:2 All Intra				YCbCr 4:2:2 Random Access			
	QP 22	**QP 27**	**QP 32**	**QP 37**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	5	3	2	1	5	4	3	2
DuckAndLegs	5	5	4	3	5	5	4	3
Kimono	5	4	2	1	5	4	3	2
OldTownCross	5	4	2	1	5	4	2	2
ParkScene	5	4	2	1	5	5	4	2
Traffic	5	4	3	2	5	5	4	2

Rounded Mean Opinion Score (Spatiotemporal Subjective Evaluation) – Pixel-PAQ versus Raw Data

Sequence	YCbCr 4:4:4 All Intra				YCbCr 4:4:4 Random Access			
	QP 22	**QP 27**	**QP 32**	**QP 37**	**QP 22**	**QP 27**	**QP 32**	**QP 37**
BirdsInCage	5	3	2	1	5	4	3	2
DuckAndLegs	5	5	4	3	5	5	4	3
Kimono	5	4	3	2	5	5	4	3
OldTownCross	5	4	2	1	5	5	4	3
ParkScene	5	4	3	2	5	4	4	3
Traffic	5	5	4	2	5	5	4	4

The MOS results tabulated in Table 5.8 indicate that there are negligible perceptual differences between the Pixel-PAQ coded versions versus the IDSQ coded versions of the following YCbCr 4:2:0 (8-bit), 4:2:2 (10-bit) and 4:4:4 (10-bit) sequences: BirdsInCage, DuckAndLegs, Kimono, OldTownCross, ParkScene and Traffic. This consistently proved to be the case for the AI and RA QP = 22 and QP = 27 tests. In most of the Pixel-PAQ versus IDSQ AI QP = 37 tests, including the experiments undertaken on the BirdsInCage 4:4:4 and 4:2:2 sequences, the chrominance quantisation-induced distortions are noticeable in the versions coded by Pixel-PAQ.

Due to the fact that BirdsInCage comprises a high level of saturation and also a combination of low variance and high variance Cb and Cr data, the coarser quantisation applied to the chroma Cb and Cr CBs by Pixel-PAQ is conspicuous in the AI QP = 37 tests. The most significant bitrate reduction achieved by Pixel-PAQ, in comparison with IDSQ, is attained on the 4:4:4 version of the BirdsInCage sequence (75% bitrate reduction in the RA QP = 22 test). Furthermore, visually lossless coding is accomplished in addition to achieving the vast bitrate reduction.

The data recorded in Table 5.9 tabulates the subjective evaluation MOS scores for the Pixel-PAQ coded sequences versus the raw video data. The goal in this set of subjective evaluations is to ascertain if Pixel-PAQ successfully achieves visually lossless coding. The subjective evaluation participants confirm that visually lossless coding is accomplished by Pixel-PAQ in the AI QP = 22 and RA QP = 22 tests in the vast majority of cases and, in some cases, in the AI QP = 27 and the RA QP = 27 tests. Very similar to the empirical findings recorded in Chapter 3 and Chapter 4 in relation to the TCPQ, CCCPQ and CBPQ techniques, the perceptual quality of the sequences coded by Pixel-PAQ using the GOP-based (inter coding) Random Access encoding configuration is considerably superior to those coded using the All Intra (intra-only) encoding configuration. This is especially true according to the tests performed with the higher initial QP values (i.e., QPs 32 and 37). To reiterate from previous chapters in this thesis, the main reason for this is because inter-coding employs advanced algorithms including motion estimation, motion compensation, advanced motion vector prediction and bidirectional inter prediction [10, 11]. Coding with I-frames only does not take into account the temporal redundancies that exist between frames, which is a well known drawback on intra-only compression. Conversely, motion data with GOP-based inter coding in HEVC can be signalled to the decoder with the utilisation of merge mode or by motion vector differences, picture reference indices and the direction of the inter prediction [10, 11]. Therefore, in the sequences coded by Pixel-PAQ, the quantisation-induced compression artifacts proved to be vastly more conspicuous in all AI tests, especially so at initial QP = 37. As such, based on the subjective evaluation results, it can be inferred that GOP-based inter coding is considerably more effective than intra-only coding in JND-based lossy video coding applications (and also lossy video coding applications in general).

5.4 Summary

A novel CB-level, JND-based luma and chroma perceptual quantisation technique is proposed for HEVC (named Pixel-PAQ). Pixel-PAQ exploits HVS-based perceptual masking, whereby spatial CSF-based luminance masking and chrominance masking are employed to achieve JND-based perceptual quantisation and visually lossless coding. The QPs for the Y CB, the Cb CB and the Cr CB are perceptually increased in order to considerably reduce bitrates without incurring a conspicuous impact on the reconstruction quality of the compressed video data. In the subjective and objective evaluations, Pixel-PAQ is compared with a state-of-the-art JND-based perceptual quantisation technique based on luminance masking (named IDSQ). In comparison with IDSQ, Pixel-PAQ achieves vast bitrate reductions — of up to 75% (an average of 72.7% over five QP data points) — on the YCbCr 4:4:4 10-bit version of the BirdsInCage sequence. In addition to this, the subjective evaluations confirmed that visually lossless coding is achieved in almost all cases in which the initial QP = 22 (for both the AI and RA tests). This proved to be the case for the Pixel-PAQ versus IDSQ tests and also the Pixel-PAQ versus raw video data tests. Finally, no significant differences in encoding and decoding times are observed for Pixel-PAQ versus IDSQ; Pixel-PAQ achieved marginal runtime reductions in all tests.

Chapter 6. Conclusions and Future Work

6.1 PhD Thesis: Overview of Contributions to Knowledge

In this thesis, four novel perceptual quantisation contributions have been proposed. In the following sub-sections, the contributions to knowledge — in addition to the conceptual novelty inherent in the proposed techniques — are summarised.

6.1.1 TCPQ (Chapter 3)

1. TCPQ [21] applies coarse perceptual quantisation to high frequency AC coefficients in the DCT frequency domain by exploiting the MTF characteristics of the HVS. By utilising a novel distance parameter, TCPQ measures the distance of an AC coefficient from the DC coefficient and subsequently discards the perceptually insignificant AC coefficients from the high frequency sub-band in the frequency domain. Compared with the ubiquitous coefficient-level quantisation technique known as RDOQ [35, 36], TCPQ attains visually lossless coding at very low bitrates.

2. To the best of the author's knowledge, TCPQ is the only perceptual quantisation technique that does not utilise spatiotemporal visual masking functions to modify the QStep. Furthermore, the subjective evaluations conducted on TCPQ provide convincing evidence that conceptually complex visual masking-based methods are not necessarily required to achieve visually lossless coding at very low bitrates.

3. The experimental evaluation results, in both the objective and subjective evaluations, highlight the fact that TCPQ is vastly more effective on high bit depth and full chroma sampling video data (i.e., YCbCr 4:4:4 10-bit data). With relevance to perceptual quantisation applications, this observation gave rise to new lines of research as regards the potential importance of chroma sampling and the bit depth of raw YCbCr data.

6.1.2 CCCPQ and CBPQ (Chapter 4)

1. CCCPQ [22] is a perceptual quantisation technique which modifies the QP at the $2N \times 2N$ CU level. It exploits the well-known HVS phenomena of luma and chroma spatial masking in high variance (busy) regions of picture data. In CCCPQ, the CU-level QP is adjusted according to the combined variances of the data in the Y, Cb and Cr CBs within a CU. Because the chroma Cb and Cr channels in raw YCbCr data contain considerable psychovisual redundancies, it is desirable to account for the variances in all three CBs in a CU when making a CU-level QP selection. This cross-colour channel dependency for QP selection, at the CU level, is a novel scheme for perceptual quantisation. In essence, based on the spatial characteristics of the data in all three colour channels, we have proved that it is possible to perceptually adjust the QP at the CU level to achieve visually lossless coding. To this end, compared with AdaptiveQP [39]-[41], which is a luminance-only perceptual quantisation technique in HEVC, CCCPQ achieves visually lossless coding at low bitrates.

2. CBPQ [23] can be considered as an extension to the proposed CCCPQ technique. However, instead of operating at the CU level, CBPQ operates at the CB level, whereby perceptual quantisation is performed on individual CBs (i.e., the Y CB, the Cb CB and the Cr CB). That is, the spatial variances of the Y CB, the Cb CB and the Cr CB are separately computed. This equates to a more refined approach in terms of separately adjusting the QP at the CB level. As is the case with the CU-level cross-colour channel dependency for QP selection aspect of CCCPQ, CB-level QP adjustment is also a novel approach to perceptual quantisation. Moreover, CBPQ exploits the CB-level QP offset signalling mechanism [18, 19], with which chroma Cb and Cr QPs can be offset against the Y QP. This allows CB-level quantisation information to be efficiently signalled to the decoder in the PPS. Note that CBPQ is tailored to high bit depth 4:4:4 video data [23]. In terms of perceptually quantising high bit depth 4:4:4 video data, CBPQ proved to be more effective than CCCPQ as regards achieving visually lossless coding at low bitrates.

3. In the CCCPQ and CBPQ tests, the experimental results show that visually lossless coding at low bitrates can be achieved by perceptually quantising chroma Cb and Cr data — particularly in a high bit depth YCbCr 4:4:4 sequence — can be quantised more aggressively. Therefore, the CCCPQ and CBPQ contributions have engendered potential new lines of research, particularly CB-level perceptual quantisation and the perceptual quantisation of chroma Cb and Cr data. Research that emerged from the CBPQ technique, for example, is the novel JND-based perceptual quantisation technique proposed in Chapter 5; i.e., Pixel-PAQ [24, 25].

6.1.3 Pixel-PAQ (Chapter 5)

1. Pixel-PAQ [24, 25] is a JND-based and CB-level perceptual quantisation technique. It exploits both luminance adaption and chrominance adaptation in order to perceptually increase the QSteps at the level of the Y CB, the Cb CB and the Cr CB. In terms of the luminance masking aspect of Pixel-PAQ, this is founded upon a previously proposed JND-based perceptual quantisation technique for HEVC that is designed for 8-bit luminance data only. However, Pixel-PAQ has significantly improved upon it by expanding compatibility to all colour channels and also YCbCr data of any bit depth.

2. Pixel-PAQ includes a novel JND-based chrominance adaptation function, which is utilised to facilitate high levels of chroma perceptual quantisation. This function takes advantage of the fact that the HVS is much less sensitive to quantisation-induced distortions in compressed chroma Cb and Cr data. Consequently, much higher levels of quantisation can be applied to data in the chroma channels without causing conspicuous distortions in the compressed video data. The aforementioned chrominance masking function included in Pixel-PAQ constitutes a novel approach to JND-based perceptual quantisation. By applying very high levels of perceptual quantisation to chroma data, tremendous bitrate reductions can be achieved, particularly when applied to high bit depth YCbCr 4:4:4 data [24]. In one experiment (an RA QP = 22 test), a 75% bitrate reduction was attained by Pixel-PAQ (compared with the reference technique).

3. This is the first JND-based perceptual quantisation technique that includes the following characteristics:

   - JND-based luminance masking and chrominance masking;
   - CB-level perceptual quantisation of Y, Cb and Cr data;
   - Compatibility with YCbCr data of any bit depth;
   - Compatibility with YCbCr data of any sampling ratio;
   - Visually lossless coding attained at extremely low bitrates;
   - Thoroughly evaluated (subjective and objective visual quality evaluations).

6.1.4 Other Contributions to Knowledge

The experimental evaluation procedure carried out on the proposed techniques can be considered as a contribution to knowledge. The novelty of the evaluation consists of the following components: i) the amalgamation of objective and subjective visual quality evaluations; ii) each contribution is evaluated on the YCbCr 4:4:4 (10-bit), 4:2:2 (10-bit) and 4:2:0 (8-bit) versions of the same sequence. Furthermore, it is interesting to note that all of the proposed perceptual quantisation techniques consistently produced the best results (in terms of bitrate reductions achieved compared with the reference techniques) on the YCbCr 4:4:4 10-bit version of each sequence.

6.1.5 Comparisons of the Proposed Techniques

In this section, we compare the bitrate reductions attained by each novel technique proposed in this thesis (i.e., TCPQ, CCCPQ, CBPQ and Pixel-PAQ). The objective here is to establish the comparative efficacy of each contribution — in comparison with Naccari's and Mrak's IDSQ technique in [42] — when applied to YCbCr 4:4:4 10-bit video data. Note that IDSQ is the reference technique with which all of the proposed methods are compared. This is because it proved to be the best performing reference method (i.e., IDSQ outperformed both RDOQ and AdaptiveQP in all tests). The following plots display the bitrates achieved by all of the aforementioned techniques (in the AI and RA tests using initial QPs 17, 22, 27, 32 and 37).

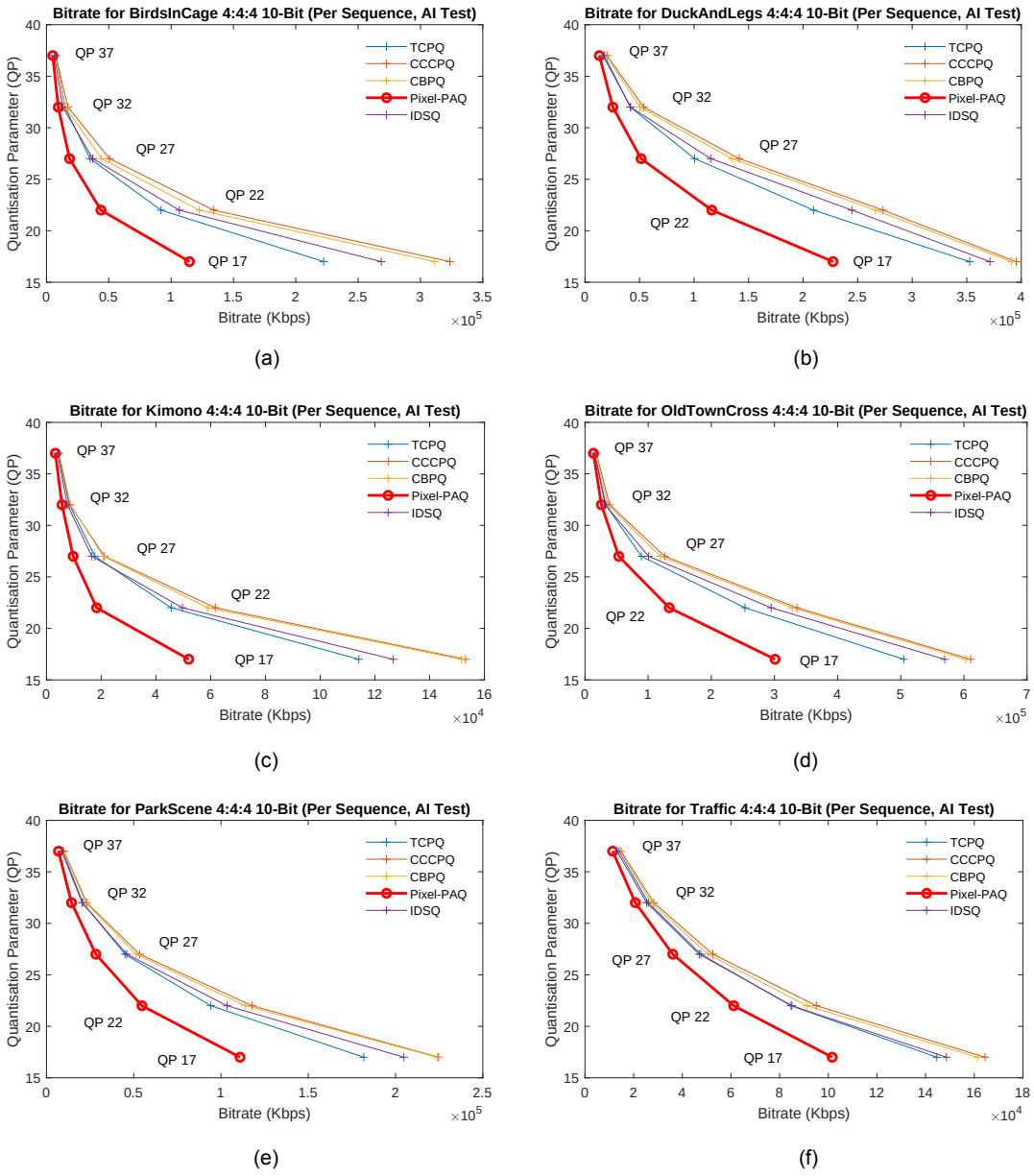


Figure 6.1: Six plots which show the comparative efficacy of TCPQ, CCCPQ, CBPQ and Pixel-PAQ (compared with IDSQ). The following subfigures show the bitrates, per sequence, per QP, in the AI tests: (a) BirdsInCage 4:4:4 10-bit; (b) DuckAndLegs 4:4:4 10-bit; (c) Kimono 4:4:4 10-bit; (d) OldTownCross 4:4:4 10-bit; (e) ParkScene 4:4:4 10-bit and (f) Traffic 4:4:4 10-bit.

The plots in Figure 6.1 highlight the fact that, in the AI tests on 4:4:4 10-bit video data, Pixel-PAQ consistently outperforms all other techniques to a considerable degree. TCPQ is the second best performing technique. CCCPQ and CBPQ are the least effective techniques in terms of bitrate reductions attained. However, it is important to note that CCCPQ and CBPQ are important to this thesis. This is because they catalysed the research for the CB-level perceptual quantisation of chrominance data; i.e., the proposed JND-based Pixel-PAQ technique emerged from this research.

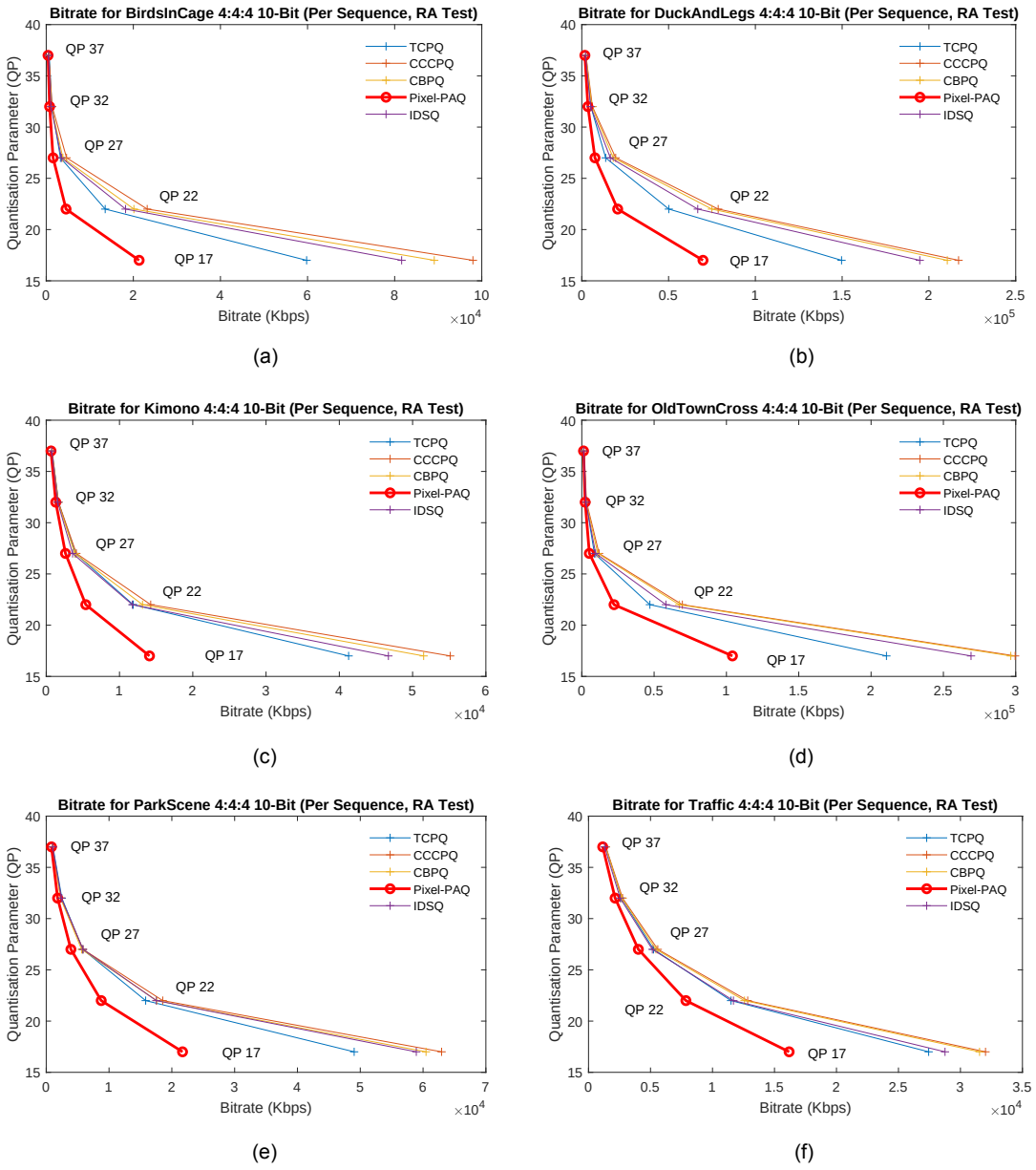


Figure 6.2: Six plots which show the comparative efficacy of TCPQ, CCCPQ, CBPQ and Pixel-PAQ (compared with IDSQ). The following subfigures show the bitrates, per sequence, per QP, in the RA tests: (a) BirdsInCage 4:4:4 10-bit; (b) DuckAndLegs 4:4:4 10-bit; (c) Kimono 4:4:4 10-bit; (d) OldTownCross 4:4:4 10-bit; (e) ParkScene 4:4:4 10-bit and (f) Traffic 4:4:4 10-bit.

As expected, the plots in Figure 6.2 mirror the pattern of results in the AI tests (as shown in the plots in Figure 6.1). In all of the RA tests conducted on 4:4:4 10-bit video data, Pixel-PAQ considerably outperforms all other techniques in terms of bitrate reductions achieved. The only difference is that greater bitrate reductions are attained by all of the proposed techniques. This is due to the fact that GOP-based spatiotemporal inter prediction — which includes motion estimation and motion compensation — is intrinsic to the RA tests.

6.2 Future Research Directions

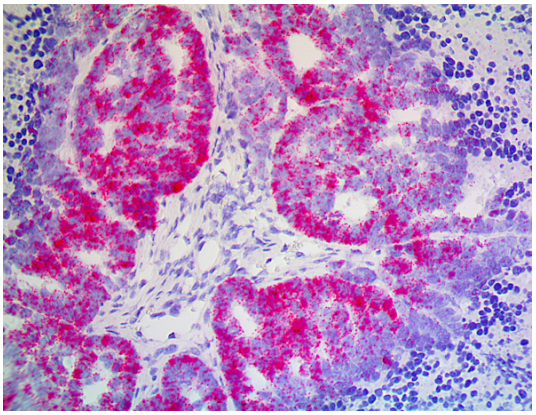


Figure 6.3: A typical whole slide image often utilised in digital medical pathology applications.

With regard to medical imaging research for which Pixel-PAQ may prove to be useful, whole slide pathology imaging, in particular, is a suitable application. As shown in Figure 6.3, whole slide images make heavy use of high levels of saturation in the colour channels of raw image data. JPEG 2000 [94] lossless coding is often perceived as the gold standard for compressing full colour whole slide medical images. JPEG 2000 lossy coding (PSNR $\geq$ 40 dB) has been previously explored for compressing full colour whole slide medical images [95, 96]. Furthermore, the HEVC standard, in both lossy mode and lossless mode, has also been utilised [97, 98]. Moreover, visually lossless coding techniques for medical imaging applications have been previously proposed for JPEG 2000; i.e., for perceptually compressing greyscale medical images [99, 56].

As previously mentioned, Pixel-PAQ is designed to discard HVS-related psychovisual redundancies — to a considerable degree — from each colour channel in YCbCr data. Therefore, Pixel-PAQ is capable of attaining vast bitrate reductions without inducing perceptually conspicuous quantisation-induced artifacts in the compressed data. In terms of achieving bitrate reductions (whilst at the same time attaining visually lossless coding), it is reasonable to postulate that Pixel-PAQ may significantly outperform previously proposed state-of-the-art medical image coding techniques, such as those proposed in [95]-[98].

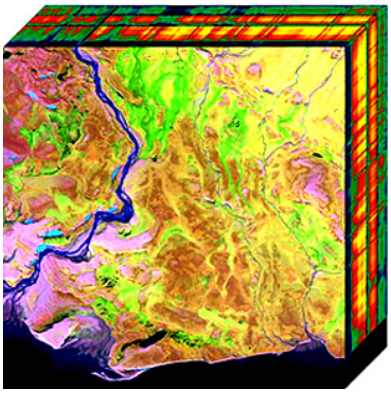


Figure 6.4: A 2D projection of a hyperspectral cube image.

Hyperspectral imaging is another interesting application for which Pixel-PAQ could potentially be utilised. In contrast to conventional raw image data, which is based on the Young-Helmholtz theory of trichromatic colour vision (i.e., RGB and YCbCr data), raw hyperspectral images typically comprise data from the entire range of the visible light portion of the electromagnetic spectrum. That is, instead of focusing on photon wavelengths in specific ranges — i.e., 690 nm, 530 nm and 470 nm (perceived by the HVS as red, green and blue, respectively) — hyperspectral images make use of the entire visible light wavelength range (i.e., from 380 nm - 750 nm); see Figure 6.4. A wide variety of applications exist for hyperspectral image coding techniques including astronomy, chemical imaging, geoscience, mineralogy, medical imaging, molecular biology, military surveillance and physics.

In terms of hyperspectral image compression, image coding engineers have developed both lossless and lossy coding techniques [100, 101]. In [101], the authors propose a novel graph wavelet method to improve the coding efficiency of the hyperspectral data. However, the technique is not perceptually optimised for the HVS in the sense that perceptual quantisation is not employed to quantise the wavelet coefficients. To this end, Pixel-PAQ could potentially augment the technique in [101] by perceptually quantising the wavelet coefficients, which may give rise to visually lossless coding at very low bitrates. It is important to note, however, that HEVC uses DCT and DST instead of the Discrete Wavelet Transform (DWT); this would need to be taken into account. Furthermore, Pixel-PAQ is designed to be used with YCbCr data. Therefore, the raw hyperspectral data would need to be converted into the YCbCr colour space prior to coding it with Pixel-PAQ.

In essence, there are several exciting and useful applications — including Whole Slide Imaging (WSI) and Hyperspectral Imaging (HI) — for which Pixel-PAQ can be employed. In terms of the coding of WSI and HI data, raw data of this nature typically requires vast bandwidth and data storage capacities (i.e., it is Big Data). Also, because of the mission critical applications in which WSI and HI data is often utilised, it is important to maximally preserve the fidelity in these types of data. Therefore, lossless compression is often used to achieve an MSE = 0 (PSNR = ∞ dB). However, due to the relatively unimpressive bitrate reductions associated with mathematically lossless compression (and also lossy compression that is not perceptually optimised), visually lossless coding — which is perceptually optimised by definition — is significantly desirable. In conclusion, contemporary whole slide imaging and hyperspectral imaging are arguably the most exciting, important and useful technical applications to which the proposed Pixel-PAQ technique may be applied.

References

[1] H. R. Wu, A. R. Reibman, W. Lin, F. Pereira, and S. S. Hemami, "Perceptual Visual Signal Compression and Transmission," *Proc. IEEE*, vol. 101, no. 9, pp. 2025–2043, 2013.

[2] W. Gao and S. Ma, "Hot Research Topics in Video Coding and Systems," in *Advanced Video Coding Systems*, Springer, 2015, pp. 211-231.

[3] T. Berger, "Rate Distortion Theory: Mathematical Basis for Data Compression," New Jersey, USA: Pearson (Prentice Hall), 1971.

[4] ITU-T Rec. H.265/HEVC (Version 5) | ISO/IEC 23008-2, Information Technology – Coding of Audio-visual Objects, *JCT-VC (ITU-T/ISO/IEC)*, 2018.

[5] G. Sullivan, J-R. Ohm, W. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649-1668, 2012.

[6] M. Wein, "Residual Coding," in *High Efficiency Video Coding: Coding Tools and Specification*, Springer, 2015, pp. 205-227.

[7] ITU-T Rec. H.264/AVS "Advanced video coding for generic audiovisual services," *ITU-T*, 2017.

[8] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560-576, 2003.

[9] J. Lainema, F. Bossen, W-J. Han, J. Min and K. Ugur, "Intra Coding of the HEVC Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1792-1801, 2012.

[10] K. Ugur, A. Alshin, E. Alshina, F. Bossen, W-J. Han, J-H. Park and J. Lainema, "Motion Compensated Prediction and Interpolation Filter Design in H.265/HEVC," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 946-955, 2013.

[11] V. Sze, M. Budagavi and G. J. Sullivan, "Inter-Picture Prediction in HEVC," in *High Efficiency Video Coding (HEVC): Algorithms and Architectures*, Springer, 2014, pp. 113-141.

[12] M. Budagavi, A. Fuldseth, G. Bjøntegaard, V. Sze and M. Sadafale, "Core Transform Design in the High Efficiency Video Coding (HEVC) Standard," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1649-1668, 2013.

[13] V. Sze, M. Budagavi and G. J. Sullivan, "HEVC Transform and Quantization," in *High Efficiency Video Coding (HEVC): Algorithms and Architectures*, Springer, 2014, pp. 141-170.

[14] I-K. Kim, J. Min, T. Lee, W-J Han and J. Park, "Block Partitioning Structure in the HEVC Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1697-1706, 2012.

[15] M. Wein, "Coding Structures," in *High Efficiency Video Coding: Coding Tools and Specification*, Springer, 2015, pp. 101-132.

[16] T. Nguyen, P. Helle, M. Winken, B. Bross, D. Marpe, H. Schwarz, T. Wiegand, "Transform Coding Techniques in HEVC," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 978- 989, 2013.

[17] M. Wein, "Quantizer Design," in *High Efficiency Video Coding: Coding Tools and Specification*, Springer, 2015, pp. 213-214.

[18] D. Flynn, D. Marpe, M. Naccari, T. Nguyen, C. Rosewarne, K. Sharman, J. Sole and J. Xu, "Overview of the Range Extensions for the HEVC Standard: Tools, Profiles, and Performance," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 26, no. 1, pp. 4-19, 2016.

[19] D. Flynn, N. Nguyen, D. He, A. Tourapis, G. Cote and D. Singer, "REext: CU adaptive chroma QP offsets," in *JCT-VC O0044, 15th Meeting of JCT-VC*, Geneva, CH, 2013, pp. 1-4.

[20] V. Sze and M. Budagavi, "High Throughput CABAC Entropy Coding in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1778-1791, 2012.

[21] L. Prangnell, V. Sanchez and R. Vanam, "Adaptive Quantization by Soft Thresholding in HEVC," in *Proc. 31st IEEE Picture Coding Symp.*, Cairns, Queensland, Australia, 2015. pp. 35-39.

[22] L. Prangnell, M. Hernández-Cabronero and V. Sanchez, "Cross Color Channel Perceptually Adaptive Quantization for HEVC," in *Proc. 27th IEEE Data Compression Conf.*, Snowbird, Utah, USA, 2017, pp. 456.

[23] L. Prangnell, M. Hernández-Cabronero and V. Sanchez, "Coding Block Level Perceptual Video Coding for 4:4:4 Data in HEVC," in *Proc. 24th IEEE Int. Conf. Image Processing*, Beijing, China, 2017, pp. 2488-2492.

[24] L. Prangnell, "Visually Lossless Coding in HEVC: A High Bit Depth and 4:4:4 Capable JND-Based Perceptual Quantisation Technique for HEVC," *Elsevier Signal Processing: Image Communication*, vol. 63, pp. 125-140, 2018.

[25] L. Prangnell and V. Sanchez, "JND-Based Perceptual Video Coding for 4:4:4 Screen Content Data in HEVC," in *Proc. 43rd IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Calgary, Alberta, Canada, 2018. In Press.

[26] F. Bossen, "Common HM test conditions and software reference configurations," in *JCT-VC L1100, 12th Meeting of JCT-VC*, Geneva, CH, 2013, pp. 1-4.

[27] C. Rosewarne, K. Sharman and D. Flynn, "Common test conditions and software reference configurations for HEVC range extensions," in *JCT-VC P1006, 16th Meeting of JCT-VC*, San José, USA, 2014, pp. 1-10.

[28] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.

[29] Joint Collaborative Team on Video Coding (JCT-VC). JCT-VC HEVC HM Reference Software, HM 16.7. Available: http://hevc.hhi.fraunhofer.de/

[30] K. McCann, C. Rosewarne, B. Bross, M. Naccari, K. Sharman and G. J. Sullivan (Editors), "HEVC Test Model 16 (HM 16) Encoder Description," in *JCT-VC R1002, 18th Meeting of JCT-VC*, Sapporo, JP, 2014, pp. 1-59.

[31] ITU-R: Rec. P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications," *ITU-R*, 2008.

[32] N. Ahmed, T. Natarajan and K. R. Rao, "Discrete Cosine Transform," *IEEE Trans. Computers*, vol. 23, no. 1, pp. 90-93, 1974.

[33] C. Wang, S. Lee and L. Chang, "Designing JPEG Quantization Tables Based on the Human Visual System," *Elsevier Signal Processing: Image Communication*, vol. 16, pp. 501-506, 2001.

[34] V. Sze, M. Budagavi and G. J. Sullivan, “Quantization Matrix,” in *High Efficiency Video Coding (HEVC): Algorithms and Architectures*, Springer, 2014, pp. 158-159.

[35] M. Karczewicz, Y. Ye and I. Chong, “Rate Distortion Optimised Quantization,” in *VCEG-AH21 (ITU-T SG16/Q6 VCEG)*, Antalya, Turkey, 2008.

[36] J. Stankowski, C. Korzeniewski, M. Domański, T. Grajek, “Rate-distortion optimized quantization in HEVC: Performance limitations,” in *Proc. 31st IEEE Picture Coding Symp.*, Cairns, Queensland, Australia, 2015. pp. 85-89.

[37] A. N. Netravali, N. J. Holmdel and B. Prasad, “Adaptive quantization of picture signals using spatial masking,” *Proc. IEEE*, vol. 65, no. 4, pp. 536-548, 1977.

[38] D. J. Connor, R. C. Brainard, and J. O. Limb, “lntra-frame Coding for Picture Transmission,” *Proc. IEEE*, vol. 60, no. 7, pp. 779-791, 1972.

[39] M. Naccari and M. Mrak, “Perceptually Optimized Video Compression,” *Elsevier Academic Press Library in Signal Processing*, vol. 5, pp. 155-196, 2014.

[40] K. Sato, “On LBS and Quantization,” in *JCT-VC D308, 4th Meeting of JCT-VC*, Daegu, KR, 2011, pp. 1-12.

[41] K. Sato, M. Budagavi, M. Coban, H. Aoki and X. Li, “CE4: Summary report of Core Experiment on quantization," in *JCT-VC F024, 6th Meeting of JCT-VC*, Torino, IT, 2011, pp. 1-20.

[42] M. Naccari and M. Mrak, “Intensity Dependent Spatial Quantization with Application in HEVC,” *IEEE Int. Conf. Multimedia and Expo*, San Jose, CA, 2013, pp. 1-6.

[43] X. H. Zhang, W. Lin, and P. Xue, “Improved Estimation for Just-Noticeable Visual Distortion,” *Elsevier Signal Processing*, vol. 85, no. 4, pp. 795-808, 2005.

[44] G. Wang, Y. Zhang, B. Li, R. Fan and M. Zhou, “A fast and HEVC compatible perceptual video coding scheme using a transform-domain Multichannel JND model,” *Springer Multimedia Tools and Applications*, vol. 76, pp. 1-27, 2017.

[45] L. Salasnich, “The Origins of Modern Physics,” in *Quantum Physics of Light and Matter: A Modern Introduction to Photons, Atoms and Many-Body Systems*, Springer, 2014, pp. 1-20.

[46] A. Chaparro, C. F. Stromeyer, E. P. Huang, R. E. Kronauer and R. T. Eskew, Jr., “Colour is what the eye sees best,” *Nature*, vol. 361, pp. 348-350, 1993.

[47] K. R. Gegenfurtner, “Cortical Mechanisms of Colour Vision,” *Nature Neuroscience*, vol. 4, pp. 563-572, 2003.

[48] H-J. Lewerenz, “Photons in Natural and Life Sciences,” Berlin, Germany: Springer, 2012.

[49] L. Salasnich, “Second Quantization of Light,” in *Quantum Physics of Light and Matter: A Modern Introduction to Photons, Atoms and Many-Body Systems*, Springer, 2014, pp. 21-48.

[50] G. Glaeser and H. F. Paulus, “The Evolution of the Eye,” Vienna, Austria: Springer, 2014.

[51] L. O. Björn, “Photobiology,” Guangzhou, China: Springer, 2015.

[52] T. Furukawa, J. B. Hurley and S. Kawamura, “Vertebrate Photoreceptors,” Osaka, Japan: Springer, 2014.

[53] ITU-R: Rec. BT.709-6 (06/2015), “Parameter Values for the HDTV Standards for Production and International Programme Exchange,” *ITU-R,* 2015.

[54] K. Rijkse, “H.263: Video Coding for Low-Bit-Rate Communication,” *IEEE Communications Magazine.*, vol. 34, no. 12, pp. 42-45, 1996.

[55] O. Avaro, A. Eleftheriadis, C. Herpel, G. Rajan and L. Ward, “MPEG-4 Systems: Overview,” Elsevier Signal Processing: Image Communication, vol. 15, pp. 281-298, 2000.

[56] H. Oh, A. Bilgin, and M. W. Marcellin, “Visually Lossless Encoding for JPEG2000,” *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 189–201, 2013.

[57] S. S. Thomas, S. Gupta, and V. K. Subramanian, “Perceptual Video Summarization—A New Framework for Video Summarization,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 8, pp. 1790-1802, 2017.

[58] Scott Janus, “Video Compression,” in *Handbook of Visual Display Technology*, Springer, 2012, pp. 287-300.

[59] Y. Dodge, “Mean Squared Error,” in *The Concise Encyclopedia of Statistics*, Springer, 2008, pp. 337-339.

[60] A. Ortega and K. Ramchandran, “Rate-Distortion Methods for Image and Video Compression,” *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 23-50, 1998.

[61] G. J. Sullivan and T. Wiegand, “Rate-Distortion Optimization for Video Compression,” *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 74-90, 1998.

[62] A. Norkin, G. Bjøntegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson, M. Zhou, and G. Auwera, “HEVC Deblocking Filter,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 22, no. 12, pp. 1746–1754, 2012.

[63] C. Fu, E. Alshina, A. Alshin, Y. Huang, C. Chen, C. Tsai, C. Hsu, S. Lei, J. Park, and W. Han, “Sample Adaptive Offset in the HEVC Standard,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 22, no. 12, pp. 1755–1764, 2012.

[64] S. W. Cheadle and S. Zeki, “Masking within and across visual dimensions: Psychophysical evidence for perceptual segregation of color and motion,” *Visual Neuroscience*, vol. 28, no. 5, pp. 445-451, 2011.

[65] ITU, “Information Technology – Digital Compression And Coding of Continuous-Tone Still Images – Requirements and Guidelines,” (JPEG), T.81, *ITU*, 1992.

[66] S. Daly, “Subroutine for the generation of a two dimensional human visual contrast sensitivity function,” *Technical Report (233203Y)*, Kodak, 1987.

[67] M. Naccari and F. Pereira, “Advanced H.264/AVC-Based Perceptual Video Coding: Architecture, Tools, and Assessment,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 21, no. 6, pp. 766-782, 2011.

[68] A. N. Netravali and C. B. Rubinstein, “Quantization of Color Signals,” *Proc. IEEE*, vol. 65, no. 8, pp. 1177-1187, 1977.

[69] J. L. Mannos and D. J. Sakrison, “The Effects of a Visual Fidelity Criterion of the Encoding of Images,” *IEEE Trans. Information Theory*, vol. 20, no. 4, pp. 525–536, 1974.

[70] A. J. Ahumada and H. A. Peterson, “Luminance-Model-Based DCT Quantization for Color Image Compression,” *Proc. SPIE*, vol. 1666, pp. 365-374, 1992.

[71] A. B. Watson, “DCTune: A Technique for Visual Optimization of DCT Quantization Matrices for Individual Images,” *Society for Information Display Digest of Technical Papers*, vol. 24. 1993, pp. 946–949.

[72] C. H. Chou and Y. C. Li, "A Perceptually Tuned Subband Image Coder Based on the Measure of Just-Noticeable-Distortion Profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 467-476, 1995.

[73] C. H. Chou and C. W. Chen, "A Perceptually Optimized 3-D Subband Codec for Video Communication Over Wireless Channels," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 6, no. 2, pp. 143-156, 1996.

[74] X. Yang, W. S. Ling, Z. Lu, E. P. Ong, and S. Yao, "Just Noticeable Distortion Model and its Applications in Video Coding," *Elsevier Signal Processing: Image Communication*, vol. 20, no. 7, pp. 662-680, 2005.

[75] Y. Jia, W. Lin, and A. A. Kassim, "Estimating Just-Noticeable Distortion for Video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 820-829, 2006.

[76] Z. Wei and K. N. Ngan, "Spatio-Temporal Just Noticeable Distortion Profile for Grey Scale Image/Video in DCT Domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 3, pp. 337–346, 2009.

[77] Z. Chen and C. Guillemot, "Perceptually-Friendly H.264/AVC Video Coding Based on Foveated Just-Noticeable-Distortion Model," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 20, no. 6, pp. 806-819, 2010.

[78] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak and D. Bull, "High Dynamic Range Video Compression Exploiting Luminance Masking," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 26, no. 5, pp. 950-964, 2016.

[79] J. Kim, S-H. Bae, and M. Kim, "An HEVC-Compliant Perceptual Video Coding Scheme Based on JND Models for Variable Block-Sized Transform Kernels," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 25, no. 11, pp. 1786-1800, 2015.

[80] S-H. Bae, J. Kim, and M. Kim, "HEVC-Based Perceptually Adaptive Video Coding Using a DCT-Based Local Distortion Detection Probability Model," *IEEE Trans. Image Processing*, vol. 25, no. 7, pp. 3343-3357, 2016.

[81] J. Wu, L. Li, W. Dong, G. Shi, W. Lin and C-C. J. Kuo, "Enhanced Just Noticeable Difference Model for Images With Pattern Complexity," *IEEE Trans. Image Processing*, vol. 26, no. 6, pp. 2682-2693, 2017.

[82] S-H. Bae and M. Kim, "A DCT-Based Total JND Profile for Spatiotemporal and Foveated Masking Effects," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 27, no. 6, pp. 1196-1207, 2017.

[83] S-H. Jung, S. K. Mitra and D. Mukherjee, “Subband DCT: definition, analysis, and applications,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 273-286, 1996.

[84] G. M. Schuster and A. K. Katsaggelos, “Review of Lossy Video Compression,” in *Rate-Distortion Based Video Compression*, Springer, 1997, pp. 13-42.

[85] V. Sze, M. Budagavi and G. J. Sullivan, “Block Structures and Parallelism Features in HEVC,” in *High Efficiency Video Coding (HEVC): Algorithms and Architectures*, Springer, 2014, pp. 49-91.

[86] V. Sze, M. Budagavi and G. J. Sullivan, “The Picture Parameter Set (PPS),” in *High Efficiency Video Coding (HEVC): Algorithms and Architectures,* Springer, 2014, pp. 32-33.

[87] ITU-R: Rec. BT.2020-2 (10/2015), “Parameter values for ultra-high definition television systems for production and international programme exchange,” *ITU-R,* 2015.

[88] ITU-R: Rec. BT.2100-1 (06/2017), “Image parameter values for high dynamic range television for use in production and international programme exchange,” *ITU-R,* 2015.

[89] G. T. Fechner, “Elements of Psychophysics, Volume 1,” (Translated by H.E. Adler), New York: Holt, Rinehart & Winston, 1860.

[90] H. R. Wu and K. R. Rao, “Digital Video Image Quality and Perceptual Coding,” 2005, pp. 640.

[91] K. T. Mullen, “The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings,” *Journal of Physiology*, vol. 359, pp. 381-400, 1985.

[92] M. Wein, “The Hybrid Video Coding Scheme,” in *High Efficiency Video Coding: Coding Tools and Specification*, Springer, 2015, pp. 37-55.

[93] V. Sze, M. Budagavi and G. J. Sullivan, “HEVC High-Level Syntax,” in *High Efficiency Video Coding (HEVC): Algorithms and Architectures*, Springer, 2014, pp. 13-48.

[94] ITU-T, “Information technology – JPEG 2000 image coding system: Core coding system,” T.800, 2015, *ITU-T*.

[95] M. Hernandez-Cabronero, F. Auli-Llinas, V. Sanchez and J. Serra-Sagrista, "Transform Optimization for the Lossy Coding of Pathology Whole-Slide Images," in *Proc. 26th IEEE Data Compression Conf.*, Snowbird, Utah, USA, 2016, pp. 131-140.

[96] M. Hernandez-Cabronero, V. Sanchez, F. Auli-Llinas and J. Serra-Sagrista, "Fast MCT Optimization for the Compression of Whole-Slide Images," in *Proc. 23rd IEEE Int. Conf. Image Processing*, Phoenix, Arizona, USA, 2016, pp. 2370-2374.

[97] V. Sanchez, F. Auli-Llinas, R. Vanam and J. Bartrina-Rapesta, "Rate Control for Lossless Region of Interest Coding in HEVC Intra Coding with Applications to Digital Pathology Images," in *Proc. 40th IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Brisbane, Australia, 2015, pp. 1250-1254.

[98] V. Sanchez and M. Hernandez-Cabronero, "Graph-based Rate Control in Pathology Imaging with Lossless Region of Interest Coding," *IEEE Trans. Medical Imaging*, 2018 (In Press).

[99] D. Wu, D. M. Tan, M. Baird, J. DeCampo, C. White and H. R. Wu, "Perceptually Lossless Medical Image Coding.," *IEEE Trans. Med. Imaging*, vol. 25, no. 3, pp. 335-344, 2006.

[100] N. Amrani, J. Serra-Sagrista, V. Laparra, M. W. Marcellin and J. Malo, "Regression Wavelet Analysis for Lossless Coding of Remote-Sensing Data.," *IEEE Trans. Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5616-5627, 2016.

[101] J. Zeng, G. Cheung, Y. H. Chao, I. Blanes, J. Serra-Sagrista and A. Ortega, "Hyperspectral Image Coding Using Graph Wavelets," in *Proc. 24th IEEE Int. Conf. Image Processing*, Beijing, China, 2017, pp. 1672-1676.